CONTINUITY, CURVATURE, AND THE GENERAL COVARIANCE OF OPTIMAL TRANSPORTATION

YOUNG-HEON KIM AND ROBERT J. MCCANN

Abstract. Let M and $\bar{M}$ be n-dimensional manifolds equipped with suitable Borel probability measures ρ and $\bar{\rho}$. Ma, Trudinger & Wang gave sufficient conditions on a transportation cost $c \in C^4(M \times \bar{M})$ to guarantee smoothness of the optimal map pushing ρ forward to $\bar{\rho}$; the necessity of these conditions was deduced by Loeper. The present manuscript shows the form of these conditions to be largely dictated by the covariance of the question; it expresses them via non-negativity of the sectional curvature of certain null-planes in a novel but natural pseudo-Riemannian geometry which the cost c induces on the product space $M \times \bar{M}$.

Hölder continuity of optimal maps was established for rougher mass distributions by Loeper, still relying on a key result of Trudinger & Wang which required certain structure on the domains and the cost. We go on to develop this theory for mass distributions on differentiable manifolds — recovering Loeper's Riemannian examples such as the round sphere as particular cases — give a direct proof of the key result mentioned above, and revise Loeper's Hölder continuity argument to make it logically independent of all earlier works, while extending it to less restricted geometries and cost functions even for subdomains M and $\bar{M}$ of $\mathbf{R}^n$. We also give new examples of geometries satisfying the hypotheses — obtained using submersions and tensor products — and some connections to spacelike Lagrangian submanifolds in symplectic geometry.

1. Introduction

Let M and $\bar{M}$ be Borel subsets of compact separable metric spaces, in which their closures are denoted by $\mathrm{cl}\, M$ and $\mathrm{cl}\, \bar{M}$. Suppose M and $\bar{M}$ are equipped with Borel probability measures ρ and $\bar{\rho}$, and let $c : \mathrm{cl}(M \times \bar{M}) \longrightarrow \mathbf{R} \cup \{+\infty\}$ be a lower semicontinuous transportation cost defined on the product space. The optimal transportation problem of Kantorovich [22] is to find the measure $\gamma \geq 0$ on $M \times \bar{M}$ which achieves the infimum

$$W_c(\rho, \bar{\rho}) := \min_{\gamma \in \Gamma(\rho,\bar{\rho})} \int_{M \times \bar{M}} c(x, \bar{x}) d\gamma(x, \bar{x}). \tag{1.1}$$

Here $\Gamma(\rho, \bar{\rho})$ denotes the set of joint probabilities having the same left and right marginals as $\rho \otimes \bar{\rho}$. It is not hard to check that this minimum is attained; any

Date: February 5, 2008.

2000 *Mathematics Subject Classification.* 49N60, 35J70, 58E17, 90B06.

This research was supported in part by Natural Sciences and Engineering Research Council of Canada Grant 217006-03 and United States National Science Foundation Grant DMS-0354729.

minimizing measure $\gamma \in \Gamma(\rho, \bar\rho)$ is then called *optimal*. Each feasible $\gamma \in \Gamma(\rho, \bar\rho)$ can be thought of as a weighted relation pairing points x distributed like ρ with points $\bar x$ distributed like $\bar\rho$; optimality implies this pairing also minimizes the average value of the specified cost $c(x, \bar x)$ of transporting each point x to its destination $\bar x$.

The optimal transportation problem of Monge [30] amounts to finding a Borel map $F : M \longrightarrow \bar M$, and an optimal measure γ vanishing outside $\mathrm{Graph}(F) := \{(x, \bar x) \in M \times \bar M \mid \bar x = F(x)\}$. When such a map F exists, it is called an *optimal map* between ρ and $\bar\rho$; in this case, the relation γ is single-valued, so that ρ-almost every point x has a unique partner $\bar x = F(x)$, and optimality can be achieved in (1.1) without subdividing the mass at such points x between different destinations. Although Monge's problem is more subtle to solve than Kantorovich's, when M is a smooth manifold and ρ vanishes on every Lipschitz submanifold of lower dimension, a *twist* condition (**(A1)**, Definition 2.1 below) on the cost function $c(x, \bar x)$ guarantees existence and uniqueness of an optimal map F, as well as uniqueness of the optimal measure γ; see Gangbo [18], Gangbo & McCann [19], Carlier [9], Ma, Trudinger & Wang [27], or the reviews in [10] [41]. One can then ask about the smoothness of the optimal map $F : M \longrightarrow \bar M$.

For Euclidean distance squared $c(x, \bar x) = |x - \bar x|^2/2$, this regularity question was resolved using geometric ideas by Caffarelli [5] [6] [7], and also by Delanoë in the plane [14] and by Urbas in higher dimensions [36], who formulated it as an oblique boundary value problem and applied the continuity method with a priori estimates. Convexity of $\bar M \subset \mathbf{R}^n$ necessarily plays a crucial role. Delanoë investigated regularity of optimal transport with respect to Riemannian distance squared on a compact manifold [29], but completed his program only for nearly flat manifolds [15], an improvement on Cordero-Erausquin's result from the torus [11], though his criterion for nearness to flat depends on the measures ρ and $\bar\rho$. Under suitable conditions on ρ and $\bar\rho$ and domains M and $\bar M \subset \mathbf{R}^n$, Ma, Trudinger & Wang [27] [34] developed estimates and a continuity method approach to a class of cost functions $c \in C^4(M \times \bar M)$ which satisfy a mysterious structure condition comparing third and fourth derivatives. Adopting the notation defined in the following section, they express this condition in the form

$$\sum_{1 \le i,j \le n} (-c_{ij\bar k\bar l} + c_{ij\bar a} c^{\bar a b} c_{\bar k\bar l b}) c^{\bar k e} c^{\bar l f} p_i p_j q_e q_f \ge C|p|^2|q|^2 \qquad \text{if } p \perp q, \tag{1.2}$$

for some constant $C \ge 0$ and each pair of orthogonal vectors $p, q \in \mathbf{R}^n$. Here summation on $\bar a, b, e, f, \bar k, \bar l$ is implicit but the sum on i, j is written explicitly for consistency with our later notation. Loeper [26] confirmed their structure condition to be necessary for continuity of F, as well as being sufficient for its smoothness [27] [34] [35]. Loeper furthermore offered a direct argument giving an explicit Hölder exponent for F, largely avoiding the continuity method, except that it relied on central results of Delanoë, Loeper, Ma, Trudinger, and Wang, to establish key technical lemmata on compact Riemannian manifolds [15] [27] such as the sphere [17], and on Euclidean domains [34].

Although the tensorial nature of condition (1.2) was known to both Trudinger and Loeper, it was not emphasized in their manuscripts.[1] Moreover, the Euclidean structure provides a deceptive identification between vectors with covectors which may obscure the geometrical content of (1.2), by suggesting that the Euclidean orthogonality of two vectors may be relevant, rather than orthogonality of a vector and covector pair. On the other hand, the Euclidean metric plays no role in the question being addressed, namely the smoothness of maps $F : M \longrightarrow \bar{M}$ which are optimal for the cost c.

Our present purpose is first to extend the theory of Loeper (and in principle, of Ma, Trudinger & Wang) to the transportation problem set on a pair of smooth manifolds M and $\bar{M}$, by finding a manifestly covariant expression of Ma, Trudinger & Wang's structure condition (1.2), as the sectional curvature non-negativity of certain null planes in a pseudo-Riemannian metric on $M \times \bar{M}$ explored here for the first time; and second to give an elementary and direct geometrical proof of the key ingredients which Loeper requires, Theorems 3.1 and A.10 below, logically independent of the methods and results of Delanoë [15], Delanoë & Loeper [17], Ma, Trudinger & Wang [27], Trudinger & Wang [34], or their subsequent work [35]. This makes Loeper's proof of Hölder continuity of optimal maps self-contained, including maps minimizing distance squared between mass distributions whose Lebesgue densities satisfy bounds above for $\rho(\cdot)$ and below for $\bar{\rho}(\cdot)$ on the round sphere $M = \bar{M} = \mathbf{S}^n$. As a byproduct of our approach, we are able to relax various geometric hypotheses on $M, \bar{M}$ and the cost c required in previous works; some of these relaxations were also obtained simultaneously and independently by Trudinger & Wang using a different approach [35], which we learned of while this paper was still in a preliminary form [25].

An important feature of our theory is in its geometric and global nature. In combination with our results from [24], this allows us to extend the conclusions of the key Theorems 3.1 and A.10 to new settings, such as the Riemannian distance squared on the product $M = \bar{M} = \mathbf{S}^{n_1} \times \cdots \times \mathbf{S}^{n_k} \times \mathbf{R}^l$ of round spheres, or Riemannian submersions thereof. This is a genuine advantage of our work over previous approaches [34] [35] [26]; see Example 3.9 and Remark 3.10

Since terms like c-convex, c-subdifferential, and notations like $\partial^c u$ are used inconsistently through the literature, and because we wish to recast the entire conceptual framework into a pseudo-Riemannian setting, we often depart from the notation and terminology developed by Ma, Trudinger, Wang and Loeper. Instead, we have tried to make the mathematics accessible to a different readership, by choosing language intended to convey the geometrical structure of the problem and its connection to

[1] Personal communications with Neil Trudinger and Gregoire Loeper. This observation is made en passant in [33], while an assertion and proof of this fact has been included in the revised version of [26], which was communicated to Villani and recorded in [41] independently of the present work. Although Loeper was motivated to call the expression appearing in (1.2) a *cost-sectional curvature* by his discoveries in the Riemannian setting (Example 3.6), it does not seem to have been noted by previous authors that the manifest covariance of the regularity question largely dictates the form of its necessary and sufficient answer, by requiring this answer to be expressible in terms of coordinate independent quantities such as geodesics and curvature.

classical concepts in differential geometry not overly specialized to optimal transportation or fully nonlinear differential equations. This approach has the advantage of inspiring certain intuitions about the problem which are quite distinct from those manifested in the previous approaches, and has a structure somewhat reminiscent of symplectic or complex geometry. Although we were initially surprised to discover that the intrinsic geometry of optimal transportation is pseudo-Riemannian, with hindsight we explain why this must be the case, and make some connections to the theory of Lagrangian submanifolds in the concluding remarks of the paper.

The outline of this paper is as follows. In the next section we introduce the pseudo-Riemannian framework and use it to adapt the relevant concepts and structures from Ma, Trudinger & Wang's work on Euclidean domains to manifolds whose only geometric structure arises from a cost function $c : M \times \bar{M} \longrightarrow \mathbf{R} \cup \{+\infty\}$. Since Morse theory prevents a smooth cost from satisfying the desired hypothesis **(A1)** on a compact manifold, we deal from the outset with functions which may fail to be smooth on a small set — such as the cut-locus in the Riemannian setting [29] [26], see Example 3.6, or the diagonal in the reflector antenna problem [21] [43] [8], see Example 3.5. This is followed by Section 3, where we motivate and state the main theorem proved here: a version of Loeper's global geometric characterization of (1.2) which we call the double-mountain above sliding-mountain maximum principle. In the same section we illustrate how this theorem and the pseudo-Riemannian framework shed new light on a series of familiar examples from Ma, Trudinger, Wang and Loeper, including those discussed above, and new ones formed from these by quotients and tensor products of, e.g., round spheres of different sizes, in Example 3.9. Section 4 contains the proofs which relate our definitions to theirs and establish the main theorem. A level set argument is required to handle the more delicate case in which the positivity in (1.2) is not strict. A last section offers some perspective on these results and their connection to optimal transportation. We have included a series of appendices which give a complete account of Loeper's theory of Hölder continuity of optimal mappings, illustrating how our main result makes this theory self-contained, and simplifying the argument at a few points. In particular, we give a unified treatment of the Riemannian sphere and reflector antenna problems, using the fact that the mapping is continuous in the former to deduce the fact that it avoids the cut-locus [17], instead of the other way around.

It is a pleasure to thank Neil Trudinger, Xu-Jia Wang, Gregoire Loeper, and Cedric Villani for many fruitful discussions, and for providing us with preprints of their inspiring works. Useful comments were also provided by John Bland, Robert Bryant, Philippe Delanoë, Stamatis Dostoglou, Nassif Ghoussoub, Gerhard Huisken, Tom Ilmanen, Peter Michor, Truyen Nguyen, and Micah Warren. We thank Jeff Viaclovsky for helpful references, and Brendan Pass for a careful reading of the manuscript. We are grateful to the Mathematical Sciences Research Institute in Berkeley, where parts of this paper were written, and to Adrian Nachman, Jim Colliander, and the 2006-07 participants Fields Analysis Working Group, for the stimulating environment which they helped to create.

2. Pseudo-Riemannian framework

Fix manifolds M and $\bar{M}$ which, if not compact, are continuously embedded in separable metrizable spaces where their closures $\mathrm{cl}\, M$ and $\mathrm{cl}\, \bar{M}$ are compact. Equip M and $\bar{M}$ with Borel probability measures ρ and $\bar{\rho}$, and a lower semicontinuous cost function $c : \mathrm{cl}(M \times \bar{M}) \longrightarrow \mathbf{R} \cup \{+\infty\}$, and a subdomain $N \subset M \times \bar{M}$ of the product manifold. Visualize the relation N as a multivalued map and denote its inverse by $N^* := \{(\bar{x}, x) \mid (x, \bar{x}) \in N\}$. We call $\bar{N}(x) := \{\bar{x} \in \bar{M} \mid (x, \bar{x}) \in N\}$ the set of destinations *visible* from x, and $N(\bar{x}) := \{x \in M \mid (x, \bar{x}) \in N\}$ the set of sources *visible* from $\bar{x}$. We define the reflected cost $c^*(\bar{x}, x) := c(x, \bar{x})$ on $\bar{M} \times M$. In local coordinates $x^1, \ldots, x^n$ on M and $x^{\bar{1}}, \ldots, x^{\bar{n}}$ on $\bar{M}$, we use the notation such as $c_i = \partial c / \partial x^i$ and $c_{\bar{i}} = \partial c / \partial x^{\bar{i}}$ to denote the partial derivatives $Dc = (c_1, \ldots, c_n)$ and $\bar{D}c = (c_{\bar{1}}, \ldots, c_{\bar{n}})$ of the cost, and $c_{i\bar{j}} = \partial^2 c / \partial x^{\bar{j}} \partial x^i$ to denote the mixed partial derivatives, which commute with each other and form the coefficients in the $n \times n$ matrix $\bar{D}Dc$. When $c_{i\bar{j}}$ is invertible its inverse matrix will be denoted by $c^{\bar{j}k}$. The same notation is used for tensor indices, with repeated indices being summed from 1 to n (or $n + 1$ to $2n$ in the case of barred indices), unless otherwise noted.

Let $T_x M$ and $T_x^* M$ denote the tangent and cotangent spaces to M at x. Since the manifold $N \subset M \times \bar{M}$ has a product structure, its tangent and cotangent spaces split canonically: $T_{(x,\bar{x})} N = T_x M \oplus T_{\bar{x}} \bar{M}$ and $T^*_{(x,\bar{x})} N = T_x^* M \oplus T_{\bar{x}}^* \bar{M}$. For $c(x, \bar{x})$ sufficiently smooth, this canonical splitting of the one-form dc will be denoted by $dc = Dc \oplus \bar{D}c$. Similarly, although the Hessian of c is not uniquely defined until a metric has been selected on N, the cross partial derivatives $\bar{D}Dc$ at $(x, \bar{x}) \in N$ define an unambiguous linear map from vectors at $\bar{x}$ to covectors at x; the adjoint $(\bar{D}Dc)^\dagger = D\bar{D}c$ of this map takes $T_x M$ to $T_{\bar{x}}^* \bar{M}$. Thus

$$h := \frac{1}{2} \begin{pmatrix} 0 & -\bar{D}Dc \\ -D\bar{D}c & 0 \end{pmatrix} \tag{2.1}$$

gives a symmetric bilinear form on the tangent space $T_{(x,\bar{x})} N$ to the product. Let us now adapt the assumptions of Ma, Trudinger & Wang [27] [34] to manifolds:
(A0)(Smoothness) $c \in C^4(N)$.

Definition 2.1. (Twist condition) *A cost $c \in C^1(N)$ is called* twisted *if*
(A1) *for all $x \in M$ the map $\bar{x} \longrightarrow -Dc(x, \bar{x})$ from $\bar{N}(x) \subset \bar{M}$ to $T_x^* M$ is injective. If c is twisted on $N \subset M \times \bar{M}$ and $c^*(\bar{x}, x) = c(x, \bar{x})$ is twisted on $N^* = \{(\bar{x}, x) \mid (x, \bar{x}) \in N\}$ we say c is* bi-twisted.

Definition 2.2. (Non-degeneracy) *A cost $c \in C^2(N)$ is* non-degenerate *if*
(A2) *for all $(x, \bar{x}) \in N$ the linear map $\bar{D}Dc : T_{\bar{x}} \bar{M} \longrightarrow T_x^* M$ is bijective.*

Though **(A1)** will not be needed until the appendices (and must extend to $\mathrm{cl}\, N$ there), for suitable probability measures ρ and $\bar{\rho}$ on M and $\bar{M}$ the twist condition alone is enough to guarantee the Kantorovich infimum (1.1) is uniquely attained, as well as existence of an optimal map $F : M \longrightarrow \bar{M}$ [18] [9] [27], as reviewed in [41] [10]. It implies the dimension of $\bar{M}$ cannot exceed that of M, while **(A2)** forces these two dimensions to coincide. The non-degeneracy condition **(A2)** ensures the map $\bar{x} \longrightarrow -Dc(x, \bar{x})$ acts as a local diffeomorphism from $\bar{N}(x) \subset \bar{M}$ to a subset

of $T^*_x M$ (which becomes global if the cost is twisted, in which case its inverse is called the *cost-exponential* [26], Definition 4.3 below), and that $h(\cdot,\cdot)$ defined by (2.1) is a non-degenerate symmetric bilinear form on $T_{(x,\bar{x})}N$. Although h is not positive-definite, it defines a pseudo-Riemannian metric on N, which might also be denoted by $d\ell^2 = -c_{i\bar{j}}dx^i dx^{\bar{j}}$. The signature of this metric is zero, since in any choice of coordinates on M and $\bar{M}$, the eigenvalues of h come in $\pm\lambda$ pairs due to the structure (2.1); the corresponding eigenvectors are $p \oplus \bar{p}$ and $(-p) \oplus \bar{p}$ in $T_{(x,\bar{x})}N = T_x M \oplus T_{\bar{x}}\bar{M}$. Non-degeneracy ensures there are no zero eigenvalues. A vector is called *null* if $h(p \oplus \bar{p}, p \oplus \bar{p}) = 0$. A submanifold $\Sigma \subset N$ is called *null* if all its tangent vectors are null vectors, and *totally geodesic* if each geodesic curve tangent to Σ at a point is contained in Σ locally. The submanifolds $\{x\} \times \bar{N}(x)$ and $N(\bar{x}) \times \{\bar{x}\}$ are examples of null submanifolds in this geometry, and will turn out to be totally geodesic. Assuming $c \in C^4(N)$, we can use the Riemann curvature tensor $R_{i'j'k'l'}$ induced by h on N to define the sectional curvature of a two-plane $P \wedge Q$ at $(x,\bar{x}) \in N$:

$$\sec_{(x,\bar{x})} P \wedge Q = \sec^{(N,h)}_{(x,\bar{x})} P \wedge Q = \sum_{i'=1}^{2n}\sum_{j'=1}^{2n}\sum_{k'=1}^{2n}\sum_{l'=1}^{2n} R_{i'j'k'l'}P^{i'}Q^{j'}P^{k'}Q^{l'}. \tag{2.2}$$

In this geometrical framework, we reformulate the mysterious structure condition (1.2) of Ma, Trudinger & Wang [27] [34] from the Euclidean setting, which was necessary for continuity of optimal maps [26] and sufficient for regularity [27] [34]. The reader is able to recover their condition from ours by computing the Riemann curvature tensor (4.2). Note that we do not normalize our sectional curvature definition (2.2) by dividing by the customary quantity $h(P,P)h(Q,Q) - h(P,Q)^2$, since this quantity vanishes in the case of most interest to us, namely $P = p \oplus 0$ orthogonal to $Q = 0 \oplus \bar{p}$, which means $p \oplus \bar{p}$ is null.

Definition 2.3. (Regular costs and cross-curvature) *A cost* $c \in C^4(N)$ *is* weakly regular *on* N *if it is non-degenerate and for every* $(x,\bar{x}) \in N$,
(A3w) $\sec_{(x,\bar{x})}(p \oplus 0) \wedge (0 \oplus \bar{p}) \geq 0$ *for all null vectors* $p \oplus \bar{p} \in T_{(x,\bar{x})}N$.
A weakly regular cost function is strictly regular *on* N *if equality in* **(A3w)** *implies* $p = 0$ *or* $\bar{p} = 0$*, in which case we say* **(A3s)** *holds on* N*. We refer to the quantity appearing in (2.3) as the* cross-curvature*, and say a weakly regular cost — and the pseudo-metric (2.1) it induces on* N *— are* non-negatively cross-curved *if*

$$\sec_{(x,\bar{x})}(p \oplus 0) \wedge (0 \oplus \bar{p}) \geq 0 \tag{2.3}$$

holds for all $(x,\bar{x}) \in N$ *and* $p \oplus \bar{p} \in T_{(x,\bar{x})}N$*, not necessarily null. The cost* c *and geometry* (N,h) *are said to be* positively cross-curved *if, in addition, equality in (2.3) implies* $p = 0$ *or* $\bar{p} = 0$*.*

If $\mathrm{cl}\, M \subset\subset M'$ and $\mathrm{cl}\, \bar{M} \subset\subset \bar{M}'$ are contained in larger manifolds and **(A0)**, **(A2)** and **(A3s/w)** all hold on some neighbourhood $N' \subset M' \times \bar{M}'$ containing $N \subset\subset N'$ compactly, we say c is strictly/weakly regular on $\mathrm{cl}\, N$. If, in addition **(A1)** holds on N', we say c is twisted on $\mathrm{cl}\, N$.

The nullity condition on $p \oplus \bar{p}$ distinguishes weak regularity of the cost from non-negative cross-curvature: this distinction is important in Examples 3.5 and 3.9

among others; see also Trudinger & Wang [34]. Non-negative cross-curvature is in turn a weaker condition than $\sec^{(N,h)} \geq 0$, which means $\sec_{(x,\bar{x})}(p \oplus \bar{q}) \wedge (q \oplus \bar{p}) \geq 0$ for all $(x, \bar{x}) \in N$ and $p \oplus \bar{q}, q \oplus \bar{p} \in T_{(x,\bar{x})}N$. As a consequence of Lemma 4.1, and due to the special form of the pseudo-metric, $\sec^{(N,h)} \geq 0$ is equivalent to requiring non-negativity of the cross-curvature operator as a quadratic form on the vector space $T_xM \wedge T_{\bar{x}}\bar{M}$, i.e.,

$$R_{i\bar{j}k\bar{l}}(p^i p^{\bar{j}} - q^i q^{\bar{j}})(p^k p^{\bar{l}} - q^k q^{\bar{l}}) \geq 0. \tag{2.4}$$

Example 2.4 (Strictly convex boundaries). Let $\Omega \subset \mathbf{R}^{n+1}$ and $\Lambda \subset \mathbf{R}^{n+1}$ be bounded convex domains with C^2-smooth boundaries. Set $M = \partial\Omega$, $\bar{M} = \partial\Lambda$, and $c(x, \bar{x}) = |x - \bar{x}|^2/2$. We claim the pseudo-metric (2.1) is non-degenerate and that $\sec^{(N,h)} \geq 0$ on $N := \{(x, \bar{x}) \in \partial\Omega \times \partial\Lambda \mid \hat{n}_\Omega(x) \cdot \hat{n}_\Lambda(\bar{x}) > 0\}$, where $\hat{n}_\Omega(x)$ denotes the outer normal to Ω at x. Indeed, fixing $(x, \bar{x}) \in N$, parameterize M near x as a graph $X \in \mathbf{R}^n \longrightarrow (X, f(X)) \in \partial\Omega$ over the hyperplane orthogonal to $\hat{n}_\Omega(x) + \hat{n}_\Lambda(\bar{x})$, and $\bar{M}$ near $\bar{x}$ by a convex graph $\bar{X} \in \mathbf{R}^n \longrightarrow (\bar{X}, g(\bar{X}))$ over the same hyperplane. This choice of hyperplane guarantees $|\nabla f(X)| < 1$ and $|\nabla g(\bar{X})| < 1$ nearby, so in the canonical coordinates and inner product on $\mathbf{R}^n$, **(A2)**–**(A3w)** follow from a computation of Ma, Trudinger & Wang [27] which yields the cross-curvature

$$\sec^{(N,h)}_{(x,\bar{x})}(p \oplus 0) \wedge (0 \oplus \bar{p}) = (p^i f_{ik} p^k)(p^{\bar{j}} g_{\bar{j}\bar{l}} p^{\bar{l}})/(2 + 2\nabla f \cdot \nabla g) \geq 0. \tag{2.5}$$

In fact, we can also deduce the stronger conclusion $\sec^{(N,h)} \geq 0$ as in (2.4):

$$\begin{aligned}
&(2 + 2\nabla f \cdot \nabla g)\ \sec^{(N,h)}_{(x,\bar{x})}(p \oplus \bar{q}) \wedge (q \oplus \bar{p}) \\
&= \langle pD^2 fp\rangle\langle \bar{p}D^2 g\bar{p}\rangle + \langle qD^2 fq\rangle\langle \bar{q}D^2 g\bar{q}\rangle - 2\langle pD^2 fq\rangle\langle \bar{p}D^2 g\bar{q}\rangle \\
&\geq \left(\sqrt{\langle pD^2 fp\rangle\langle \bar{p}D^2 g\bar{p}\rangle} - \sqrt{\langle qD^2 fq\rangle\langle \bar{q}D^2 g\bar{q}\rangle}\right)^2.
\end{aligned} \tag{2.6}$$

Noting $\hat{n}_\Omega((X, f(X))) = (\nabla f(X), -1)$, Ma, Trudinger & Wang's computation shows nondegeneracy **(A2)** fails at the boundary of N where $\hat{n}_\Omega(x) \cdot \hat{n}_\Lambda(\bar{x}) = 0$ implies the denominator of (2.5) is zero. Gangbo & McCann [20] showed the cost is twisted on N provided Λ is strictly convex, but cannot be twisted on any larger domain in $M \times \bar{M}$. If both Ω and Λ are 2-uniformly convex, meaning that the Hessians $D^2 f$ and $D^2 g$ are positive definite, the conditions for equality in (2.6) show the sectional curvature of (N, h) to be positive. The resulting strict regularity **(A3s)** underlies Gangbo & McCann's proof of continuity for each of the multiple mappings which — due to the absence of twisting **(A1)** — are required to support the unique optimizer $\gamma \in \Gamma(\rho, \bar{\rho})$ in this geometry. Here the probability measures ρ and $\bar{\rho}$ are assumed mutually absolutely continuous with respect to surface measure on Ω and Λ, both having densities bounded away from zero and infinity. For contrast, observe in this case that the same computations show that although nondegeneracy **(A2)** also holds on the set $N_- := \{(x, \bar{x}) \in \partial\Omega \times \partial\Lambda \mid \hat{n}_\Omega(x) \cdot \hat{n}_\Lambda(\bar{x}) < 0\}$, this time both surfaces can be expressed locally as graphs over $\hat{n}_\Omega(x) - \hat{n}_\Lambda(\bar{x})$ but **(A3w)** fails at each point $(x, \bar{x}) \in N_-$: indeed, the cross-curvatures of N_- are all negative because $D^2 f > 0 > D^2 g$ have opposite signs.

Let us now exploit the geodesic structure which the pseudo-metric h induces on $N \subset M \times \bar{M}$ to recover Ma, Trudinger & Wang's notions concerning c-convex domains [27] in our setting.

Definition 2.5. (Notions of convexity) *A subset $W \subset N \subset M \times \bar{M}$ is* geodesically convex *if each pair of points in W is linked by a curve in W satisfying the geodesic equation on (N, h). This definition is extended to subsets $W \subset \operatorname{cl} N$ by allowing geodesics in N which have endpoints on ∂N. We say $\bar{\Omega} \subset \operatorname{cl} \bar{M}$* appears convex *from $x \in M$ if $\{x\} \times \bar{\Omega}$ is geodesically convex and $\bar{\Omega} \subset \operatorname{cl} \bar{N}(x)$. We say $W \subset M \times \bar{M}$ is* vertically convex *if $\bar{W}(x) := \{\bar{x} \in \bar{M} \mid (x, \bar{x}) \in W\}$ appears convex from x for each $x \in M$. We say $\Omega \subset \operatorname{cl} M$* appears convex *from $\bar{x} \in \bar{M}$ if $\Omega \times \{\bar{x}\}$ is geodesically convex and $\Omega \subset \operatorname{cl} N(\bar{x})$. We say $W \subset M \times \bar{M}$ is* horizontally convex *if $W(\bar{x}) := \{x \in M \mid (x, \bar{x}) \in W\}$ appears convex from $\bar{x}$ for each $\bar{x} \in \bar{M}$. If W is both vertically and horizontally convex, we say it is* bi-convex.

For a non-degenerate twisted cost **(A0)**–**(A2)**, Lemma 4.4 shows $\bar{\Omega} \subset \bar{N}(x)$ appears convex from x if and only if $Dc(x, \bar{\Omega})$ is convex in $T^*_x M$; similarly for a bi-twisted cost $\Omega \subset N(\bar{x})$ appears convex from $\bar{x}$ if and only if $\bar{D}c(\Omega, \bar{x})$ is convex in $T^*_{\bar{x}} \bar{M}$. This leads immediately to notions of apparent *strict* convexity, and apparent *uniform* convexity for such sets, and shows our definition of *apparent* convexity is simply an adaptation to manifolds of the c-convexity and c^*-convexity of Ma, Trudinger & Wang [27]: Ω is c^*-convex in their language with respect to $\bar{x}$ if it appears convex from $\bar{x}$; $\bar{\Omega}$ is c-convex with respect x if it appears convex from x; and Ω and $\bar{\Omega}$ are c^*- and c-convex with respect to each other if $N = \Omega \times \bar{\Omega} \subset \mathbf{R}^{2n}$ is bi-convex, meaning $Dc(x, \bar{N}(x)) \subset T^*_x M$ and $\bar{D}c(N(\bar{x}), \bar{x}) \subset T^*_{\bar{x}} \bar{M}$ are convex domains for each $(x, \bar{x}) \in N$.

Remark 2.6. In dimension $n = 1$, strict regularity **(A3s)** follows vacuously from non-degeneracy **(A2)**, since $p \oplus \bar{p}$ null implies $p = 0$ or $\bar{p} = 0$. For this reason we generally discuss $n \geq 2$ hereafter. Note however, for $n = 1$ and $c \in C^4(N)$ non-degenerately twisted, N is bi-convex if and only if $N(\bar{x})$ and $\bar{N}(x)$ are homeomorphic to intervals. In local coordinates x^1 and $\bar{x}^1$ on N, non-degeneracy implies $c_{1\bar{1}} = \mp e^{\pm \lambda(x^1, \bar{x}^1)}$. Comparing $-c_{11\bar{1}\bar{1}} + c_{11\bar{1}} c^{\bar{1}1} c_{1\bar{1}\bar{1}} = \lambda_{1\bar{1}} |c_{1\bar{1}}|$ with (4.2) shows

$$c(x^1, \bar{x}^1) = \mp \int_{x_0}^{x^1} \int_{\bar{x}_0}^{\bar{x}^1} e^{\pm \lambda(s,t)} ds dt \tag{2.7}$$

induces a pseudo-metric h on N for which $\sec_{(x,\bar{x})}(p \oplus \bar{0}) \wedge (0 \oplus \bar{p})$ has the same sign as $\partial^2 \lambda / \partial x^1 \partial \bar{x}^1$ whenever $p \neq 0 \neq \bar{p}$. If (N, h) is connected its cross-curvature will have therefore a definite sign if $\lambda(x, \bar{x})$ is non-degenerate, and a semidefinite sign if $\lambda(x, \bar{x})$ is twisted. Moreover, the sign of the cross-curvature, sectional curvature, and curvature operator all coincide on a product of one-dimensional manifolds, although this would not necessarily be true on a product of surfaces or higher-dimensional manifolds.

Definition 2.7. (c-contact set) *Given $\Omega \subset \operatorname{cl} M$, $u : \Omega \longrightarrow \mathbf{R} \cup \{+\infty\}$, and $c : \operatorname{cl}(M \times \bar{M}) \longrightarrow \mathbf{R} \cup \{+\infty\}$, we define* $\operatorname{Dom} c := \{(x, \bar{x}) \in \operatorname{cl}(M \times \bar{M}) \mid c(x, \bar{x}) < \infty\}$,

the c-contact set $\partial^c_\Omega u(x) := \{\bar x \in \mathrm{cl}\,\bar M \mid (x,\bar x) \in \partial^c_\Omega u\}$*, and* $\partial^c u = \partial^c_{\mathrm{cl}\,M} u$*, where*

$$\partial^c_\Omega u := \{(x,\bar x) \in \mathrm{Dom}\, c \mid u(y) + c(y,\bar x) \geq u(x) + c(x,\bar x) \text{ for all } y \in \Omega\}. \tag{2.8}$$

We define $\mathrm{Dom}\,\partial^c_\Omega u := \{x \in \Omega \mid \partial^c_\Omega u(x) \neq \emptyset\}$ *and* $\mathrm{Dom}\,\partial^c u := \mathrm{Dom}\,\partial^c_{\mathrm{cl}\,M} u$.

3. Main Results and Examples

A basic result of Loeper [26] states that a cost satisfying **(A0)-(A2)** on a bi-convex domain $N = M \times \bar M \subset \mathbf{R}^n \times \mathbf{R}^n$ is weakly regular **(A3w)** if and only if $\partial^c u(x)$ appears convex from x for each function $u : M \longrightarrow \mathbf{R} \cup \{+\infty\}$ and each $x \in M$. His necessity argument is elementary and direct, but for sufficiency he appeals to a result of Trudinger & Wang which required c-boundedness of the domains M and $\bar M$ in the original version of [34]. The same authors gave another proof of sufficiency for strictly regular costs in [35], and removed the c-boundedness restriction in the subsequent revision of [34]. Our main result is a direct proof of this sufficiency, found independently but simultaneously with [35], under even weaker conditions on the cost function and domain geometry. In particular, the manifolds M and $\bar M$ in Theorem 3.1 need not be equipped with global coordinate charts or Riemannian metrics, the open set $N \subset M \times \bar M$ need not have a product structure, and the weakly regular cost need neither be twisted nor strictly regular. This freedom proves useful in Examples 2.4 and 3.9 and Remark 3.10.

Theorem 3.1. (Weak regularity connects c-contact sets) *Use a cost* $c : \mathrm{cl}(M \times \bar M) \longrightarrow \mathbf{R} \cup \{+\infty\}$ *with non-degenerate restriction* $c \in C^4(N)$ *to define a pseudo-metric (2.1) on a horizontally convex domain* $N \subset M \times \bar M$*. Fix* $\Omega \subset \mathrm{cl}\,M$*,* $x \in M$*, and a set* $\bar\Omega \subset \mathrm{cl}\,\bar N(x)$ *which appears convex from* x*. Suppose* $\cap_{0 \leq t \leq 1} N(\bar x(t))$ *is dense in* Ω *for each geodesic* $t \in [0,1] \longrightarrow (x, \bar x(t)) \in \{x\} \times \bar\Omega$*, and* $c : \mathrm{cl}(N) \longrightarrow \mathbf{R} \cup \{+\infty\}$ *is continuous. If* c *is weakly regular* **(A3w)** *on* N*, then* $\bar\Omega \cap \partial^c_\Omega u(x)$ *is connected (and in fact appears convex from* x*) for each* $u : \Omega \longrightarrow \mathbf{R} \cup \{+\infty\}$ *with* $x \in \mathrm{Dom}\, u$*.*

To motivate the proof of this theorem and its relevance to the economics of transportation, consider the optimal division of mass $\rho(\cdot)$ between two target points $\bar y, \bar z \in \bar M$ in ratio $(1-\epsilon)/\epsilon$. This corresponds to the minimization (1.1) with $\bar\rho(\cdot) = (1-\epsilon)\delta_{\bar y}(\cdot) + \epsilon \delta_{\bar z}(\cdot)$. If $c(x,\bar x)$ is the cost of transporting each commodity unit from x to $\bar x$, a price differential λ between the fair market value of the same commodity at $\bar z$ and $\bar y$ will tend to balance demand ϵ and $1-\epsilon$ with supply $\rho(\cdot)$, given the relative proximity of $\bar z$ and $\bar y$ to the producers $\rho(\cdot)$ distributed throughout M. Here proximity is measured by transportation cost. Since each producer will sell his commodity at $\bar z$ or $\bar y$, depending on which of these two options maximizes his profit, the economic equilibrium and optimal solution will be given, e.g. as in [19], by finding the largest $\lambda \in \mathbf{R}$ such that

$$\begin{aligned} u(x) &= \max\{\lambda - c(x,\bar y), -c(x,\bar z)\} \\ \text{yields} \quad \epsilon &\leq \rho\left[\{x \in M \mid u(x) = -c(x,\bar z)\}\right]. \end{aligned} \tag{3.1}$$

Producers in the region $\{x \in M \mid \lambda - c(x,\bar y) > -c(x,\bar z)\}$ will choose to sell their commodities at $\bar y$, while producers in the region where the opposite inequality

holds will choose to sell their commodities at $\bar{z}$; points $x_0 \in M$ on the hypersurface $c(x_0, \bar{y}) - c(x_0, \bar{z}) = \lambda$ of equality are indifferent between the two possible sale destinations $\bar{y}$ and $\bar{z}$. We call this hypersurface the *valley of indifference*, since it corresponds to a crease in the graph of the function u. Loeper's observation is that for optimal mappings to be continuous, each point x_0 in the valley of indifference between $\bar{y}$ and $\bar{z}$ must also be indifferent to a continuous path of points $\bar{x}(t)$ linking $\bar{y} = \bar{x}(0)$ to $\bar{z} = \bar{x}(1)$; otherwise, he constructs a measure ρ concentrated near x_0 for which the optimal map to a mollified version of $\bar{\rho}$ exhibits a discontinuous jump, since arbitrarily close producers will choose to supply very different consumers. Indifference means one can choose $\lambda(t)$ such that

$$u(x) \geq \max_{0<t<1} \lambda(t) - c(x, \bar{x}(t)) \tag{3.2}$$

for all $x \in M$, with equality at x_0 for each $t \in [0,1]$. When the path connecting $\bar{y}$ to $\bar{z}$ exists, this equality forces $\lambda(t) = c(x_0, \bar{x}(t)) - c(x_0, \bar{z})$; it also forces the path $t \in [0,1] \longrightarrow (x_0, \bar{x}(t))$ to be a geodesic for the pseudo-metric (2.1) on (N, h), so the path $\{\bar{x}(t) \mid 0 \leq t \leq 1\}$ appears convex from x_0.

We think of a function of the form $x \longrightarrow \lambda - c(x, \bar{y})$ as defining the elevation of a mountain on M, *focused at* (or indexed by) $\bar{y} \in \bar{M}$. The function $u(x)$ of (3.1) may be viewed as a *double mountain*, while the maximum (3.2) may be viewed as the upper envelope of a one-parameter family of mountains which slide as their foci $\bar{x}(t)$ move from $\bar{y}$ to $\bar{z}$. The proof of the preceding theorem relies on the fact that the sliding mountain stays beneath the double mountain (while remaining tangent to it at x_0), if the cost is weakly regular. In the applications below, we take $\Omega = \mathrm{cl}\, M$ and $\bar{\Omega} = \mathrm{cl}\, \bar{N}(x) = \mathrm{cl}\, \bar{M}$ tacitly.

Proof of Theorem 3.1. Let $c \in C^4(N)$ be weakly regular on some horizontally convex domain $N \subset M \times \bar{M}$. Fix $u : \Omega \longrightarrow \mathbf{R} \cap \{+\infty\}$ and $x \in \mathrm{Dom}\, u$ with $\bar{y}, \bar{z} \in \bar{\Omega} \cap \partial^c_\Omega u(x)$. This means $u(y) \geq u(x) - c(y, \bar{z}) + c(x, \bar{z})$ for all $y \in \Omega$, the right hand side takes an unambiguous value in $\mathbf{R} \cup \{-\infty\}$, and the same inequality holds with $\bar{y}$ in place of $\bar{z}$. Apparent convexity of $\bar{\Omega}$ from x implies there exists a geodesic $t \in]0,1[\longrightarrow (x, \bar{x}(t))$ in (N, h) with $\bar{x}(t) \in \bar{\Omega}$ which extends continuously to $\bar{x}(0) = \bar{y}$ and $\bar{x}(1) = \bar{z}$. The desired connectivity can be established by proving $\bar{x}(t) \in \partial^c_\Omega u(x)$ for each $t \in]0,1[$, since this means $\bar{\Omega} \cap \partial^c_\Omega u(x)$ appears convex from x.

Horizontal convexity implies $N(\bar{x}(t))$ appears convex from $\bar{x}(t)$ for each $t \in [0,1]$, so agrees with the illuminated set $V(x, \bar{x}(t)) = N(\bar{x}(t))$ of Definition 4.7. For any $y \in \cap_{0 \leq t \leq 1} N(\bar{x}(t))$, the sliding mountain lies below the double mountain, i.e., $f(t, y) := -c(y, \bar{x}(t)) + c(x, \bar{x}(t)) \leq \max\{f(0^+, y), f(1^-, y)\}$, according to Theorem 4.10 and Remark 4.12. Note that $f : [0,1] \times \Omega \longrightarrow \mathbf{R} \cup \{-\infty\}$ is a continuous function, since $t \in [0,1] \to (x, \bar{x}(t)) \in \mathrm{cl}\, N$ and $c : \mathrm{cl}(N) \longrightarrow \mathbf{R} \cup \{+\infty\}$ are continuous and their composition is real-valued. We therefore replace 0^+ by 0 and 1^- by 1 and extend the inequality to all $y \in \Omega$ using the density of $\cap_{0 \leq t \leq 1} N(\bar{x}(t))$. On the other hand, $\bar{x}(t) \in \partial^c_\Omega u(x)$ holds if and only if $u(y) \geq u(x) + f(t, y)$ for each $y \in \Omega$ and $t \in [0,1]$. Since $u(y) \geq u(x) + \max\{f(0, y), f(1, y)\}$ by hypothesis,

we have established apparent convexity of $\bar{\Omega} \cap \partial^c_\Omega u(x)$ from x and the proof is complete. $\square$

Remark 3.2 (Hölder continuity). Since connectedness and apparent convexity survive closure, we may replace $\bar{\Omega}\cap\partial^c_\Omega u(x)$ by its closure (often $\partial^c u(x)$) without spoiling the result. The apparent convexity of $\partial^c u(x)$ from x hints at a kind of monotonicity for the correspondence $x \in M \longrightarrow \partial^c u(x)$. A strict form of this monotonicity can be established when the cost is strictly regular (**A3s**) (Proposition B.1), and was exploited by Loeper to prove Hölder continuity $F \in C^{1/(4n-1)}(M; \mathrm{cl}\,\bar{M})$ of the optimal map between densities $\rho \in L^\infty(M)$ and $\bar\rho$ with $(1/\bar\rho) \in L^\infty(M)$ for costs which are strictly regular and bi-twisted on the closure of a bi-convex domain $M \times \bar{M} \subset\subset \mathbf{R}^{2n}$. Details of his argument and conclusions are given in the appendices below.

Remark 3.3 (On the relevance of twist and apparent convexity to the converse). In the absence of the twist condition, we have defined apparent convexity by the existence of a geodesic, which need not be extremal or unique. When **(A3w)** fails in this general setting, Loeper's converse argument shows the existence of a geodesic segment with endpoints in $(\{x\} \times \bar{N}(x)) \cap \partial^c u$ but which departs from this set at some points in between. Since the twist condition and apparent convexity imply the existence and uniqueness of geodesics linking points in $\{x\} \times \bar{N}(x)$, for a twisted cost the *if* statement in Theorem 3.1 becomes necessary as well as sufficient, a possibility which was partly anticipated in Ma, Trudinger & Wang [27].

Remark 3.4 (Product Domains). If $N = M \times \bar{M}$ the hypotheses and conclusions become simpler to state because $N(\bar{x}) = M$ and $\bar{N}(x) = \bar{M}$ for each $(x, \bar{x}) \in N$. If, in addition the product $N = M \times \bar{M} \subset \mathbf{R}^{2n}$ is a bounded Euclidean domain, we recover the result proved by Loeper [26] based on the regularity results of Trudinger & Wang [34], whose hypotheses were relaxed after [35].

Example 3.5 (The reflector antenna and conformal geometry). The restriction of the cost function $c(x, \bar{x}) = -\log|x - \bar{x}|$ from $\mathbf{R}^n \times \mathbf{R}^n$ to the unit sphere $M = \bar{M} = \mathbf{S}^{n-1} := \partial\mathbf{B}^n_1(0)$ arises in conjunction [43] [21] with the reflector antenna problem studied by Caffarelli, Glimm, Guan, Gutierrez, Huang, Kochengin, Marder, Newman, Oliker, Waltmann, Wang, and Wescott, among others. It induces a metric h known to satisfy **(A0)-(A3s)** with $N = (M \times \bar{M}) \setminus \Delta$. Note however, that $c(x, \bar{x})$ actually defines a pseudo-metric $h^{(\infty,\infty)}$ on the larger space $\mathbf{R}^n \times \mathbf{R}^n \setminus \Delta$ which is almost but not quite bi-convex. Here $\Delta := \{(y, y) \mid y \in \mathbf{R}^n \cup \{\infty\}\}$ denotes the diagonal. For fixed $a \neq \bar{a} \in \mathbf{R}^n$,

$$\tilde{c}^{(a,\bar{a})}(x, \bar{x}) := -\frac{1}{2} \log \frac{|x - \bar{x}|^2 |a - \bar{a}|^2}{|x - \bar{a}|^2 |a - \bar{x}|^2}$$

induces a pseudo-metric $h^{(a,\bar{a})}$ which coincides with $h^{(\infty,\infty)}$ on the set where both can be defined. Moreover, $\tilde{c}^{(a,\bar{a})}$ extends smoothly to $(\tilde{M}_a \times \tilde{M}_{\bar{a}}) \setminus \Delta$ where $\tilde{M} = \mathbf{R}^n \cup \{\infty\}$ is the Riemann sphere and $\tilde{M}_a = \tilde{M} \setminus \{a\}$. Furthermore $h^{(a,\bar{a})}$ is independent of $(a, \bar{a})$, so has a (unique) extension $\tilde{h}$ to $\tilde{N} := (\tilde{M} \times \tilde{M}) \setminus \Delta$ which turns out to satisfy **(A2)-(A3s)** on the bi-convex set $\tilde{N}$; as in [26] [27] [34], this can be verified using

the alternate characterization of **(A3w/s)** via concavity (/ 2-uniform concavity) of the restriction of the function

$$q^* \in T_x^* \tilde{M} \longrightarrow p^i p^j c_{ij}(x, c\text{-Exp}_x q^*) \tag{3.3}$$

to the nullspace of $p \in T_x \tilde{M}$ in $\mathrm{Dom}\, c\text{-Exp}_x$; here the c-Exp map is defined at (4.3). The nullspace condition is crucial, since the value of this function is given by

$$p^i p^j f_{ij}|_{(Df)^{-1}(-q^*)} = 2(q_i^* p^i)^2 - |p|^2 |q^*|^2, \tag{3.4}$$

where the left-hand expression in (3.4) coincides with the right-hand expression in (3.3) for general costs of the form $c(x, \bar{x}) = f(x - \bar{x})$. Homogeneity of the resulting manifold $(\tilde{N}, \tilde{h})$ follows from the symmetries $\tilde{N} = (F \times F)(\tilde{N})$ and

$$\tilde{c}^{(a,\bar{a})}(x, \bar{x}) = \tilde{c}^{(F(a), F(\bar{a}))}(F(x), F(\bar{x}))$$

under simultaneous translation $F(x) = x - y$ by $y \in \mathbf{R}^n$ or inversion $F(x) = x/|x|^2$ of both factor manifolds; note the identity $|x - \bar{x}| = |x||\bar{x}|||x|^{-2}x - |\bar{x}|^{-2}\bar{x}|$. This simplifies the verification of **(A2)-(A3s)**, since it means infinity plays no distinguished role. Bi-convexity of $\tilde{N}$ follows from the fact that the projection of the null geodesic through $(z, \bar{x})$ and $(y, \bar{x})$ onto M is given by a portion of the circle in $\mathbf{R}^n \cup \{\infty\}$ passing through z, y and $\bar{x}$ — the unique arc of this circle (or line) stretching from z to y which does not pass through $\bar{x}$. From homogeneity it suffices to compute this geodesic in the case $(z, \bar{x}) = (\infty, 0)$; we may further take $y = (1, 0, \ldots, 0)$ using invariance under simultaneous rotations $F(x) = \Lambda x$ by $\Lambda \in O(n)$ and dilations $F(x) = \lambda x$ by $\lambda > 0$. This calculation demonstrates that $\cap_{0 \le t \le 1} \tilde{N}(\bar{x}(t))$ is the complement of a circular arc — hence a dense subset of $\tilde{M}$ — for each geodesic $t \in [0,1] \longrightarrow (x, \bar{x}(t)) \in \tilde{N}$. Notice that if $x, \bar{x}(0), \bar{x}(1)$ all lie on the sphere $M = \bar{M}$, then so does $\bar{x}(t)$ for each $t \in [0,1]$. After verifying that h coincides with the restriction of $\tilde{h}$ to the codimension-2 submanifold N, we infer for $\bar{x} \in \partial \mathbf{B}_1^n(0)$ that $N(\bar{x}) \times \{\bar{x}\}$ is a totally geodesic hypersurface in $\tilde{N}(\bar{x}) \times \{\bar{x}\}$, so the horizontal and vertical geodesics, bi-convexity, and strict regularity of (N, h) are inherited directly from the geometry of $(\tilde{N}, \tilde{h})$, the strict regularity via Lemma 4.5. Although we lack a globally defined smooth cost on $\tilde{N}$, we have one on N, so the hypotheses and hence the conclusions of Theorems 3.1 and 4.10 are directly established in the reflector antenna problem, whose geometry is also clarified: the vertical geodesics are products of points on M with circles on $\bar{M}$, where by circle we mean the intersection of a two-dimensional plane with $\bar{M} = \partial \mathbf{B}_1^n(0)$. The geodesics would be the same for the negation $c_+(x, y) = +\log|x - y|$ of this cost, which satisfies **(A0)-(A2)** on the bi-convex domain N, but violates **(A3w)** for the same reason that c satisfies **(A3s)**. Using the Euclidean norm on our coordinates, the c-exponential (4.3) is given by $\bar{x} = c\text{-Exp}_x(q^*) := x - q^*/|q^*|^2$, the optimal map takes the form $F(x) = c\text{-Exp}_x Du(x)$, and the resulting degenerate elliptic Monge-Ampère type equation (5.1)–(5.2) on $\mathbf{R}^n$, expressed in coordinates, is

$$\det[u_{ij}(x) + 2u_i(x)u_j(x) - \delta_{ij}|Du(x)|^2] = \frac{|Du(x)|^{2n}\rho(x)}{\bar{\rho}(x - Du(x)/|Du(x)|^2)}. \tag{3.5}$$

Here $\bar\rho(\bar x) = O(|\bar x|^{-2n})$ as $\bar x \to \infty$ since $\bar\rho$ transforms like an n-form under coordinate changes. Using the isometries above to make $Du(x_0) = 0$ at a point x_0 of interest, a slight perturbation of the standard Monge-Ampère equation is recovered nearby. The operator under the determinant (3.5) is proportional to the Schouten tensor of a conformally flat metric $ds^2 = e^{-4u}\sum_{i=1}^n (dx^i)^2$, so a similar equation occurs in Viaclovsky's σ_n-version of the Yamabe problem [38] [39], which has been studied by the many authors in conformal geometry surveyed in [37] and [33].

Example 3.6 (Riemannian manifolds). Consider a Riemannian manifold $(M = \bar M, g)$. Taking the cost function to be the square of the geodesic distance $c(x,\bar x) = d^2(x,\bar x)/2$ associated to g, induces a pseudo-metric tensor (2.1) on the domain N where $c(x,\bar x)$ is smooth, i.e. the complement of the cut locus. Moreover, the cost exponential (Definition 4.3) reduces [29] to the Riemannian exponential

$$c\text{-Exp}_x p^* = \exp_x p \tag{3.6}$$

with the metrical identification $p^* = g(p,\cdot)$ of tangent and cotangent space. A curve $t \in [0,1] \longrightarrow x(t) \in N(\bar x)$ through $\bar x$ is a geodesic in (M,g) if and only if the curve $\tau(t) = (x(t),\bar x)$ is a (null) geodesic in (N,h), according to Lemma 4.4. On the diagonal $x = \bar x$, we compute $h(p\oplus\bar p, p\oplus\bar p) = g(p,\bar p)$, meaning the pseudo-Riemannian space (N,h) contains an isometric copy of the Riemannian space (M,g) along its diagonal $\Delta := \{(x,x) \mid x \in M\}$. The symmetry $c(x,\bar x) = c(\bar x, x)$ shows Δ to be embedded in N as a totally geodesic submanifold, and nullity of $p \oplus \bar p \in T_{(x,x)}N$ reduces to orthogonality of p with $\bar p$. This perspective illuminates Loeper's observation [26] that negativity of one Riemannian sectional curvature at any point on (M,g) violates weak regularity **(A3w)** of the cost. Indeed, the comparison of (4.9) with Lemma 4.5 allows us to recover the fact that along the diagonal, cross-curvatures in (N,h) are proportional to Riemannian curvatures in (M,g):

$$\sec^{(N,h)}_{(x,x)}(p\oplus 0)\wedge(0\oplus\bar p) = \frac{4}{3}\sec^{(M,g)}_x p\wedge\bar p. \tag{3.7}$$

Example 3.7 (The round sphere). In the case of the sphere $M = \bar M = \mathbf{S}^n$ equipped with the standard round metric, the cut locus consists of pairs $(x,\bar x)$ of antipodal points $d(x,\bar x) = \operatorname{Diam} M$. Denote its complement by $N = \{(x,\bar x) \mid d(x,\bar x) < \operatorname{Diam} M\}$, where $c(x,\bar x) = d^2(x,\bar x)/2$ is smooth. In this case, the identification (3.6) of cost exponential with Riemannian exponential implies bi-convexity of N, since the cut locus forms a circle (hypersphere if $n > 2$) in the tangent space $T_x\mathbf{S}^n$, and the verification of (**A3s**) both on and off the diagonal was carried out by Loeper [26]. In fact more is true: (N,h) is non-negatively cross-curved, as we verify in [24]. Given an h-geodesic $t \in [0,1] \longrightarrow (x,\bar x(t)) \in N$, we find $\cap_{0\le t\le 1} N(\bar x(t))$ exhausts $\mathbf{S}^n$ except for the antipodal curve to the exponential image $t \in [0,1] \longrightarrow \bar x(t)$ of a line segment in $T_x\mathbf{S}^n$. Although this curve does not generally lie on a great circle or even a circle, its complement is dense in $\mathbf{S}^n$, whence the double-mountain above sliding-mountain property and Theorem 3.1 follow.

Example 3.8 (Positive versus negative curvature). To provide more insight into Loeper's results — continuity of optimal maps on the round sphere versus discontinuous optimal maps on the saddle or hyperbolic plane — consider dividing a smooth

positive density ρ (say uniform on some disk of volume $\sum_{i=1}^3 \epsilon_i = 1$) optimally between three points $\bar{x}_1, \bar{x}_2, \bar{x}_3$ on a geodesic through its center: $\bar{\rho}(\cdot) = \sum_{i=1}^3 \epsilon_i \delta_{\bar{x}_i}(\cdot)$. The solution to this problem is given, e.g. [19], by finding constants $\lambda_1, \lambda_2, \lambda_3 \in \mathbf{R}$ for which the function

$$u(x) = \max_{1 \le i \le 3} \{u_i(x)\} \quad \text{given by} \quad u_i(x) := -\lambda_i - c(x, \bar{x}_i) \tag{3.8}$$
$$\text{solves } \epsilon_i = \rho[\Omega_i] \qquad \text{with} \qquad \Omega_i := \{x \in M \mid u(x) = u_i(x)\}$$

for $i = 1, 2, 3$. The regions Ω_i are illustrated in Figure 1 for the case where cost $c(x, \bar{x}) = d^2(x, \bar{x})/2$ is proportional to Riemannian distance squared on the (a) round sphere, (b) Euclidean plane, (c) hyperbolic disc.

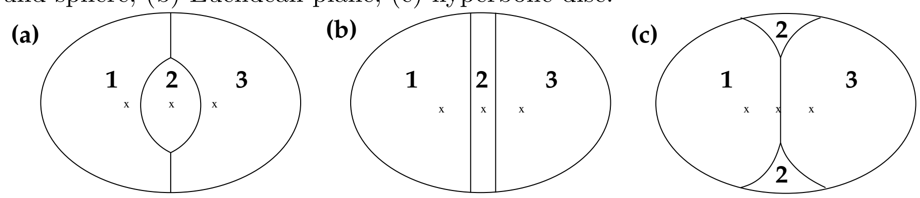

Figure 1

Notice that only in the third case is one of the regions Ω_2 disconnected. Now consider transporting a smoothly smeared and positive approximation $\bar{\rho}_\eta(\cdot)$ of the discrete source $\bar{\rho}(\cdot)$, obtained by mollifying on scale $\eta > 0$, to the uniform measure $\rho(\cdot)$. For $\eta > 0$ sufficiently small, points $\bar{x}$ near $\bar{x}_2$ will need to map into a δ-neighbourhood of Ω_2. It is not hard to believe the results of Loeper's calculation for case (c), namely that if $\eta > 0$ is small the map from $\bar{\rho}_\eta(\cdot)$ to $\rho(\cdot)$ will have a discontinuity near $\bar{x}_2$ separating the regions which map into disjoint δ-neighbourhoods separating the two disconnected components of Ω_2.

Example 3.9 (New examples from old: perturbations, submersions, and products). It remains interesting to find more general sufficient conditions on a Riemannian manifold (M, g) and function f for the pseudo-metric h induced on the complement of the cut locus $N \subset M \times M$ by $c(x, \bar{x}) = f(d(x, \bar{x}))$ to be strictly or weakly regular. It is clear that slight perturbations of an **(A3s)** cost remain strictly regular if the perturbation is smooth and small enough. Delanoë & Ge have quantified this observation on the round sphere [16]. It is also possible to deduce that the local properties such as **(A3s/w)** and non-negative cross-curvature and the global property such as the conclusion of Theorem 3.1 all survive Riemannian submersion [24], holding for example on quotients of the round sphere under discrete [16] [24] or continuous [24] group actions, including in particular all spaces of constant positive curvature [16], and the Fubini-Study metric on $\mathbf{CP}^n$ [24]. On the other hand, an example constructed by one of us shows that positive but non-constant sectional curvature of the underlying manifold $(M = \bar{M}, g)$ does not guarantee weak regularity of the cost $c = d^2/2$ away from the diagonal in $N \subset M \times M$ [23]. As a final important example, consider two manifolds $N_+ \subset M_+ \times \bar{M}_+$ and $N_- \subset M_- \times \bar{M}_-$ equipped with cost functions $c_\pm \in C^4(N_\pm)$ inducing pseudo-metrics $d\ell_\pm^2 = -c_{i\bar{j}} dx_\pm^i dx_\pm^{\bar{j}}$. As a consequence of Lemma 4.5, the product metric $d\ell^2 = d\ell_+^2 + d\ell_-^2$ corresponding

to the cost function $c_+(x_+, \bar{x}_+) + c_-(x_-, \bar{x}_-)$ on $N = (M_+ \times M_-) \times (\bar{M}_+ \times \bar{M}_-)$ is non-negatively cross-curved (2.3) if both $(N_\pm, h_\pm)$ are. Although it is not true that weak regularity of the factors implies the same for the product, many of the known examples of weakly regular costs (including those of Examples 2.4, 3.7 and the sumbersions above) actually turn out to be non-negatively cross-curved [24], so this product construction becomes a fruitful source of new examples. Furthermore, since geodesics in the product are products of geodesics, bi-convexity of the factors $(N_\pm, h_\pm)$ implies bi-convexity of the product manifold (N, h). Because a product geodesic may have constant factors, it is not hard to show that the non-negatively curved product manifolds (N, h) always fail to be strictly regular even when both factor manifolds $(N_\pm, h_\pm)$ are positively cross-curved. Thus tensor products of positively cross-curved costs yield a new source of weakly regular costs that fail to be strictly regular — the very simplest example of which is given by arbitrary sums $c(x, \bar{x}) = \sum_{k=1}^n c(x^k, x^{\bar{k}}; k)$ of $k = 1, \ldots, n$ positively cross-curved costs $c(s, t; k)$ as in (2.7) on bi-convex subdomains $N_1, \ldots, N_n \subset \mathbf{R}^2$.

Remark 3.10. (Products $\mathbf{S}^{n_1} \times \cdots \times \mathbf{S}^{n_k} \times \mathbf{R}^l$ and their Riemannian submersions) The conclusion of Theorem 3.1 holds for the distance squared cost on the Riemannian product $M = \bar{M}$ of round spheres $\mathbf{S}^{n_1} \times \cdots \times \mathbf{S}^{n_k} \times \mathbf{R}^l$ — or its Riemannian submersions — by combining the preceding example with the result of [24]. The weak regularity **(A3w)** of the cost on $N = M \times \bar{M} \setminus \{\text{cut locus}\}$ and the biconvexity of N are satisfied as in Example 3.9. The density condition of $\cap_{0 \le t \le 1} N(\bar{x}(t))$ follows also easily since the cut locus of one point in this example is a smooth submanifold of codimension greater than or equal to 2. This new global result illustrates an advantage of our method over other approaches [34] [35] [26], which would require a regularity result for optimal maps (or some a priori estimates) to obtain the conclusion of Theorem 3.1. To implement such approaches for the manifolds of this example, one would need to establish that an optimal map remains *uniformly away from the cut locus*, as is currently known only for a single sphere $M = \bar{M} = \mathbf{S}^n$ from work of Delanöe & Loeper [17] (alternately [26] or Corollary E.2). To the best of our knowledge, no one has yet succeeded in establishing regularity results for this product example, though one could try to obtain them from Theorems 3.1 and A.10, by extending the approach we develop successfully in Appendix E for a single sphere $M = \bar{M} = \mathbf{S}^n$.

4. Proof of Main Results

Let us begin by establishing coordinate representations of the Christoffel symbols and Riemann curvature tensor for the pseudo-metric (2.1).

Lemma 4.1. (Riemann curvature tensor and Christoffel symbols) *Use a non-degenerate cost $c \in C^4(N)$ to define a pseudo-metric (2.1) on the domain $N \subset M \times \bar{M}$. In local coordinates $x^1, \ldots, x^n$ on M and $x^{\bar{1}}, \ldots, x^{\bar{n}}$ on $\bar{M}$, the only non-vanishing Christoffel symbols are*

$$\Gamma_{ij}{}^{m} = c^{m\bar{k}} c_{\bar{k}ij} \qquad \text{and} \qquad \Gamma_{\bar{i}\bar{j}}{}^{\bar{m}} = c^{\bar{m}k} c_{k\bar{i}\bar{j}}. \tag{4.1}$$

Furthermore, the components of the Riemann curvature tensor (2.2) vanish except when the number of barred and unbarred indices is equal, in which case the value of the component can be inferred from $R_{ij\bar{k}\bar{l}} = 0$ *and*

$$2R_{i\bar{j}\bar{k}l} = c_{i\bar{j}\bar{k}l} - c_{li\bar{f}} c^{a\bar{f}} c_{a\bar{j}\bar{k}}. \tag{4.2}$$

Proof. From

$$\Gamma_{ij}{}^{m} := \frac{1}{2} h^{mk}(h_{kj,i} + h_{ik,j} - h_{ij,k}) + \frac{1}{2} h^{m\bar{k}}(h_{\bar{k}j,i} + h_{i\bar{k},j} - h_{ij,\bar{k}}),$$

and the analogous definition with i, j, and/or m replaced by $\bar{i}, \bar{j}$ and $\bar{m}$ respectively, the off-diagonal form (2.1) of the pseudo-metric and the equality of mixed partials implies the only non-vanishing Christoffel symbols are given by

$$\begin{aligned} \Gamma_{ij}{}^{m} &= -\frac{1}{2} c^{m\bar{k}}(-c_{\bar{k}ji} - c_{i\bar{k}j} + 0) \\ &= c^{m\bar{k}} c_{\bar{k}ij} \end{aligned}$$

and $\Gamma^{\bar{k}}_{\bar{i}\bar{j}} = c^{\bar{m}k} c_{k\bar{i}\bar{j}}$. Since the only nonvanishing Christoffel symbols are given by (4.1), it is not hard to compute the relevant components of Riemann's curvature tensor in the coordinates we have chosen:

$$\begin{aligned} R_{i\bar{j}\bar{k}}{}^{\bar{m}} &= -\frac{\partial}{\partial x^{i}} \Gamma^{\bar{m}}_{\bar{j}\bar{k}} + \frac{\partial}{\partial x^{\bar{j}}} \Gamma^{\bar{m}}_{i\bar{k}} + \Gamma^{f}_{\bar{k}i} \Gamma^{\bar{m}}_{\bar{j}f} - \Gamma^{f}_{\bar{k}\bar{j}} \Gamma^{\bar{m}}_{if} + \Gamma^{\bar{f}}_{\bar{k}i} \Gamma^{\bar{m}}_{\bar{j}\bar{f}} - \Gamma^{\bar{f}}_{\bar{k}\bar{j}} \Gamma^{\bar{m}}_{i\bar{f}} \\ &= -\frac{\partial}{\partial x^{i}} (c^{\bar{m}a} c_{a\bar{j}\bar{k}}) \\ &= -c^{\bar{m}b} c_{i\bar{j}\bar{k}b} + c^{\bar{m}b} c_{bi\bar{f}} c^{a\bar{f}} c_{a\bar{j}\bar{k}}. \end{aligned}$$

Using the pseudo-metric (2.1) to lower indices yields (4.2), and the other non-vanishing components of the Riemann tensor can then be deduced from the well-known symmetries $-R_{\bar{j}i\bar{k}l} = R_{i\bar{j}\bar{k}l} = R_{\bar{k}li\bar{j}} = -R_{l\bar{k}i\bar{j}}$. To see that the remaining components all vanish, it suffices to repeat the analysis starting from the analogous definitions of $R_{ij\bar{k}}{}^{\bar{m}}$, $R_{ij\bar{k}}{}^{m}$, and $R_{ijk}{}^{\bar{m}}$. □

Remark 4.2 (Vanishing curvatures). The vanishing of R_{ijkl}, $R_{ijk\bar{l}}$, $R_{i\bar{j}\bar{k}\bar{l}}$, $R_{\bar{i}\bar{j}\bar{k}\bar{l}}$, and $R_{ij\bar{k}\bar{l}} = 0$ imply that $(\wedge^2 TM) \oplus (\wedge^2 T\bar{M})$ lies in the null space of curvature operator viewed as a quadratic form on $\wedge^2 TN = (\wedge^2 TM) \oplus (\wedge^2 T\bar{M}) \oplus (TM \wedge T\bar{M})$. Strict / weak regularity of the cost, and the signs of the cross-curvature (2.3), and sectional curvature (2.4) are therefore all determined by the action of this operator on the n^2 dimensional vector bundle $TM \wedge T\bar{M}$. Since the cone of null vectors is nonlinear, it is not obvious whether **(A3w)** implies non-negativity of all cross-curvatures of (N, h), but Trudinger & Wang [34] prove, as in Example 3.5, that this is not so.

We next recall an important map of Ma, Trudinger & Wang [27], called the *cost-exponential* by Loeper [26].

Definition 4.3. (Cost exponential) *If* $c \in C^2(N)$ *is twisted* **(A1)***, we define the* c-exponential *on*

$$\begin{aligned} \mathrm{Dom}(c\text{-}Exp_x) :=& -Dc(x, \bar{N}(x)) \\ =& \{p^* \in T^*_x M \mid p^* = -Dc(x, \bar{x}) \quad \textit{for some } \bar{x} \in \bar{N}(x)\} \end{aligned} \tag{4.3}$$

by c-$Exp_x p^* = \bar{x}$ if $p^* = -Dc(x,\bar{x})$. Non-degeneracy **(A2)** *then implies the c-exponential is a diffeomorphism from* $\mathrm{Dom}(c\text{-}Exp_x) \subset T^*_x M$ *onto* $\bar{N}(x) \subset \bar{M}$*. If* $c \in C^2(N)$ *is non-degenerate but not twisted and* $q^* = -Dc(x,\bar{y})$*, the implicit function theorem implies a single-valued branch of c-*Exp_x *taking values near* $\bar{y}$ *is defined by the same formula in a small neighbourhood of* q^**, though it no longer extends to be a global diffeomorphism of* $\mathrm{Dom}(c\text{-}Exp_x)$ *onto* $\bar{N}(x)$.

Lemma 4.4. (the c-segments of [27] are geodesics) *Use a non-degenerate cost* $c \in C^4(N)$ *to define a pseudo-metric (2.1) on the domain* $N \subset M \times \bar{M}$*. Fix* $x \in M$*. If* $p^*, q^* \in \mathrm{Dom}(c\text{-}Exp_x) \subset T^*_x M$ *are close enough together there will be a branch of c-*Exp_x *defined on the line segment joining* p^* *to* q^**. Then the curve* $s \in [0,1] \longrightarrow \sigma(s) := (x, c\text{-}Exp_x((1-s)p^* + sq^*))$ *is an affinely parameterized null geodesic in (N,h). Conversely, every geodesic segment in the totally geodesic submanifold* $\{x\} \times \bar{N}(x)$ *can be parameterized locally in this way.*

Proof. Given $s_0 \in [0,1]$, introduce coordinates on M and $\bar{M}$ around $\sigma(s_0)$ so that nearby, the curve $\sigma(s)$ can be represented in the form $(x^1, \ldots, x^n, x^{\bar{1}}(s), \ldots, x^{\bar{n}}(s))$. Differentiating the definition of the cost exponential

$$0 = (1-s)p^*_i + sq^*_i + c_i(\sigma(s)) \tag{4.4}$$

twice with respect to s yields

$$0 = c_{i\bar{j}}\ddot{x}^{\bar{k}} + c_{i\bar{j}\bar{k}}\dot{x}^{\bar{j}}\dot{x}^{\bar{k}} \tag{4.5}$$

for each $i = 1, \ldots, n$. Multiplying by the inverse matrix $c^{\bar{m}i}$ to $c_{i\bar{j}}$ yields

$$0 = \ddot{x}^{\bar{m}} + c^{\bar{m}i}c_{i\bar{j}\bar{k}}\dot{x}^{\bar{j}}\dot{x}^{\bar{k}}, \tag{4.6}$$

always summing on repeated indices. Together with $\ddot{x}^m = 0 = \dot{x}^m$, we claim these reduce to the geodesic equations

$$\begin{aligned} 0 &= \ddot{x}^m + \Gamma_{ij}{}^m\dot{x}^i\dot{x}^j + \Gamma_{\bar{i}j}{}^m\dot{x}^{\bar{i}}\dot{x}^j + \Gamma_{i\bar{j}}{}^m\dot{x}^i\dot{x}^{\bar{j}} + \Gamma_{\bar{i}\bar{j}}{}^m\dot{x}^{\bar{i}}\dot{x}^{\bar{j}} \\ 0 &= \ddot{x}^{\bar{m}} + \Gamma_{ij}{}^{\bar{m}}\dot{x}^i\dot{x}^j + \Gamma_{\bar{i}j}{}^{\bar{m}}\dot{x}^{\bar{i}}\dot{x}^j + \Gamma_{i\bar{j}}{}^{\bar{m}}\dot{x}^i\dot{x}^{\bar{j}} + \Gamma_{\bar{i}\bar{j}}{}^{\bar{m}}\dot{x}^{\bar{i}}\dot{x}^{\bar{j}} \end{aligned} \tag{4.7}$$

on (N,h). Indeed, this follows since the only non-vanishing Christoffel symbols are given by (4.1). Comparing (4.6) with (4.7), we see $\sigma(s)$ is an affinely parameterized geodesic near $\sigma(s_0)$; it is null since $h(\dot{\sigma},\dot{\sigma}) = 0$ from the off-diagonal form of (2.1).

Conversely, any geodesic segment in (N,h) which lies in $x \times \bar{N}(x)$ can be parameterized affinely on $s \in [0,1]$. Near $s_0 \in [0,1]$ it then satisfies (4.5), whence

$$0 = \frac{d^2}{ds^2}c_i(x,\bar{x}(s)). \tag{4.8}$$

Integrating twice, the constants of integration determine p^* and $q^* \in T^*_x M$ such that (4.4) holds locally. Thus $(1-s_0)p^* + s_0 q^* \in \mathrm{Dom}(c\text{-}\mathrm{Exp}_x)$. Choosing a branch of the cost exponential defined near this point and equalling $D_i c(x,\bar{x}(s_0))$ there, we deduce $(x,\bar{x}(s)) = c\text{-}\mathrm{Exp}_x((1-s)p^* + sq^*)$ for s near s_0 from the definition of this branch.

Finally, to see that $\{x\} \times \bar{N}(x)$ is totally geodesic, take any point $\bar{x} \in \bar{N}(x)$ and tangent vector $\bar{p} \in T_{\bar{x}}\bar{M}$. Setting $x^m(s) = x^m$ to be constant solves half of the geodesic equations, since $\Gamma^m_{\bar{i}j} = 0 = \Gamma^m_{\bar{i}\bar{j}}$. we can still solve the remaining n

components of the geodesic equation (4.6) for small $s \in \mathbf{R}$, subject to the initial conditions $\bar{x}(0) = \bar{x}$ and $\dot{\bar{x}}(0) = q$, to find a geodesic which remains in the n-dimensional submanifold $\{x\} \times \bar{N}(x)$ for short times. $\square$

The next lemma gives a non-tensorial expression of the sectional curvature in our pseudo-Riemannian geometry (N, h). In the context of Example 3.6, it can be viewed as a generalization of the asymptotic formula for the Riemannian distance between two arclength parameterized geodesics $x(s)$ and $\bar{x}(t)$ near a point of intersection $x(0) = \bar{x}(0)$ at angle θ:

$$d^2(x(s), \bar{x}(t)) = s^2 + t^2 - 2st\cos\theta - \frac{k}{3}s^2t^2\sin^2\theta + O((s^2+t^2)^{5/2}) \tag{4.9}$$

where the Riemannian curvature k of the 2-plane $\dot{x}(0) \wedge \dot{\bar{x}}(0)$ on $(M = \bar{M}, g)$ gives the leading order correction to the law of cosines. Though we do not need it here, the proof of the next lemma can also be adapted to establish an expansion analogous to (4.9) for general costs $c(x(s), \bar{x}(t))$; the zeroth and first order terms do not vanish, but the coefficients of s^2t and s^3t are zero due to the geodesy of $s \in [0,1] \longrightarrow \sigma(s)$. Remarkably however, to determine the coefficient of s^2t^2 in the lemma below requires only one (in fact, either one) and not both of the two curves to be geodesic.

Lemma 4.5. (Non-tensorial expression for curvature) *Use a non-degenerate cost $c \in C^4(N)$ to define a pseudo-metric (2.1) on the domain $N \subset M \times \bar{M}$. Let $(s,t) \in [-1,1]^2 \longrightarrow (x(s), \bar{x}(t)) \in N$ be a surface containing two curves $\sigma(s) = (x(s), \bar{x}(0))$ and $\tau(t) = (x(0), \bar{x}(t))$ through $(x(0), \bar{x}(0))$. Note $0 \oplus \dot{\bar{x}}(0)$ defines a parallel vector-field along $\sigma(s)$. If $s \in [-1,1] \longrightarrow \sigma(s) \in N$ is a geodesic in (N,h) then*

$$-2\frac{\partial^4}{\partial s^2 \partial t^2}\bigg|_{s=0=t} c(x(s), \bar{x}(t)) = \sec_{(x(0),\bar{x}(0))} \frac{d\sigma}{ds} \wedge \frac{d\tau}{dt}. \tag{4.10}$$

Proof. Introduce coordinates $x^1, \ldots, x^n$ in a neighbourhood of $x(0)$ on M and $x^{\bar{1}}, \ldots, x^{\bar{n}}$ in a neighbourhood of $\bar{x}(0)$ on $\bar{M}$, so the surface $(x(s), \bar{x}(t)) \in N$ has coordinates $(x^1(s), \ldots, x^n(s), x^{\bar{1}}(t), \ldots, x^{\bar{n}}(t))$ locally. To see $0 \oplus \dot{\bar{x}}(0)$ defines a parallel vector field along $\sigma(s)$, we use the Levi-Civita connection to compute

$$\dot{\sigma}^i \nabla_i \dot{x}^{\bar{k}} + \dot{\sigma}^{\bar{i}} \nabla_{\bar{i}} \dot{x}^{\bar{k}} = \dot{x}^i \frac{\partial \dot{x}^{\bar{k}}}{\partial x^i} + \dot{x}^i \Gamma^{\bar{k}}_{ij} \dot{x}^j + \dot{x}^i \Gamma^{\bar{k}}_{i\bar{j}} \dot{x}^{\bar{j}} + 0 = 0$$

since the only non-vanishing Christoffel symbols are given by (4.1).

Computing the fourth mixed derivative yields

$$\begin{aligned} &\frac{\partial^4}{\partial s^2 \partial t^2}\bigg|_{s=0=t} c(x(s), \bar{x}(t)) \\ = \quad & c_{ij\bar{k}\bar{l}} \dot{x}^i \dot{x}^j \dot{x}^{\bar{k}} \dot{x}^{\bar{l}} + c_{a\bar{k}\bar{l}} \ddot{x}^a \dot{x}^{\bar{k}} \dot{x}^{\bar{l}} + (c_{ij\bar{b}} \dot{x}^i \dot{x}^j + c_{a\bar{b}} \ddot{x}^a) \ddot{x}^{\bar{b}} \\ = \quad & (c_{i\bar{k}\bar{l}j} - c_{\bar{k}\bar{l}a} c^{a\bar{b}} c_{\bar{b}ij}) \dot{x}^i \dot{x}^j \dot{x}^{\bar{k}} \dot{x}^{\bar{l}} \end{aligned} \tag{4.11}$$

where the form (4.6) of the geodesic equation has been used to eliminate the coefficient of $\ddot{x}^{\bar{b}}$ and express $\ddot{x}^a$ in terms of $\dot{x}^i$. Comparing (4.11) with (4.2) and (2.2)

yields the desired conclusion (4.10). The minus sign comes from antisymmetry $R_{i\bar{k}\bar{l}j} = -R_{i\bar{k}j\bar{l}}$ of the Riemann tensor. □

Our next contribution culminates in Theorem 4.10, which generalizes the result that Loeper [26] deduced from Trudinger & Wang [34]. As mentioned above, it can be interpreted to mean that if a weakly regular function $c \in C^4(N)$ governs the cost of transporting a commodity from the locations where it is produced to the locations where it is consumed, a shipper indifferent between transporting the commodity from y to the consumer at either endpoint $\bar{x}(0)$ and $\bar{x}(1)$ of the geodesic $t \in [0,1] \longrightarrow (y, \bar{x}(t))$ in (N, h), will also be indifferent to transporting goods from y to the consumers at each of the intermediate points $\bar{x}(t)$ along this geodesic. As was also mentioned, for non-degenerate costs Loeper showed this conclusion fails unless the cost is weakly regular.

Proposition 4.6. **(Maximum principle)** *Use a weakly regular cost $c \in C^4(N)$ to define a pseudo-metric (2.1) on the domain $N \subset M \times \bar{M}$. Given $x \neq y \in M$, let $t \in]0,1[\longrightarrow (x, \bar{x}(t)) \in N$ be a geodesic in (N, h) with $\dot{\bar{x}}(1/2) \neq 0$ and set $f(t) = -c(y, \bar{x}(t)) + c(x, \bar{x}(t))$. If $\dot{f}(t_0) = 0$ for some $t_0 \in]0,1[$ with a geodesic linking $(x, \bar{x}(t_0))$ to $(y, \bar{x}(t_0))$ lying in $N(\bar{x}(t_0)) \times \{\bar{x}(t_0)\}$, then $\ddot{f}(t_0) \geq 0$. Strict inequality holds if the relevant cross-curvature of c is positive at some point on the second geodesic.*

Proof. Suppose $f(t)$ has a critical point at some $0 < t_0 < 1$, so

$$0 = \dot{f}(t_0) = \left(\partial_{\bar{i}} c(x, \bar{x}(t_0)) - \partial_{\bar{i}} c(y, \bar{x}(t_0))\right) \dot{x}^{\bar{i}}(t_0); \tag{4.12}$$

we then claim $\ddot{f}(t_0) \geq 0$.

Let $s \in [0,1] \longrightarrow (x(s), \bar{x}(t_0))$ be a geodesic in $N(\bar{x}(t_0)) \times \{\bar{x}(t_0)\}$ with endpoints $x(0) = x$ and $x(1) = y$. Setting $g(s,t) = -c(x(s), \bar{x}(t)) + c(x, \bar{x}(t))$, Lemma 4.5 yields

$$\left.\frac{\partial^4 g}{\partial s^2 \partial t^2}\right|_{(s,t_0)} = \frac{1}{2} \sec_{(x(s),\bar{x}(t_0))} (\dot{x}(s) \oplus 0) \wedge (0 \oplus \dot{\bar{x}}(t_0)) \geq 0, \tag{4.13}$$

with the inequality following from weak regularity **(A3w)** of the cost $c \in C^4(N)$, as long as $\dot{x}(s) \oplus 0_{\bar{x}(t_0)}$ and $0_{x(s)} \oplus \dot{\bar{x}}(t_0)$ are orthogonal vectors on (N,h), or equivalently, as long as $\dot{x}(s) \oplus \dot{\bar{x}}(t_0)$ is null. These vectors are non-vanishing since $x(0) \neq x(1)$ and $\dot{\bar{x}}(0) \neq 0$; we now deduce their orthogonality from (4.12), using subscripts to distinguish which tangent space the zero vectors reside in.

Along the geodesic $s \in [0,1] \longrightarrow (x(s), \bar{x}(t_0))$, the vector field $0_{x(s)} \oplus \dot{\bar{x}}(t_0)$ is parallel transported according to Lemma 4.5. Thus the inner product λ of this vector field with the tangent vector is independent of $s \in [0,1]$. Define $q^*_{\bar{i}}(s) := \partial_{\bar{i}} c(x(s), \bar{x}(t_0)) \in T^*_{\bar{x}(t_0)} \bar{M}$, so $\dot{q}^*_{\bar{i}}(s) = \partial_j \partial_{\bar{i}} c(x(s), \bar{x}(t_0)) \dot{x}^j(s)$. From the form (2.1) of the pseudo-metric we discover

$$\begin{aligned} \lambda &= h(0_{x(s)} \oplus \dot{\bar{x}}(t_0), \dot{x}(s) \oplus 0_{\bar{x}(t_0)}) \\ &= -\dot{x}^{\bar{i}}(t_0) \dot{q}^*_{\bar{i}}(s). \end{aligned}$$

Integrating this constant over $0 < s < 1$, (4.12) yields the desired orthogonality

$$\lambda = \dot{x}^{\bar{i}}(t_0)(q^*_{\bar{i}}(0) - q^*_{\bar{i}}(1)) = 0.$$

Now (4.13) shows $\partial^2 g/\partial t^2|_{t=t_0}$ to be a convex function of $s \in [0,1]$. We shall prove this convex function is minimized at $s = 0$, where it vanishes. Introducing coordinates $x^1, \ldots, x^n$ around $x = x(0)$ on M and $x^{\bar{1}}, \ldots, x^{\bar{n}}$ around $\bar{x}(t_0)$ on $\bar{M}$, we compute

$$\begin{aligned} \frac{\partial^2 g}{\partial t^2}\bigg|_{(s,t_0)} &= -\Big[c_{\bar{i}}(x(s),\bar{x}(t))\ddot{x}^{\bar{i}} + c_{\bar{i}\bar{j}}(x(s),\bar{x}(t))\dot{x}^{\bar{i}}\dot{x}^{\bar{j}}\Big]^{(s,t_0)}_{(0,t_0)} \\ \frac{\partial^3 g}{\partial s \partial t^2}\bigg|_{(s,t_0)} &= -(c_{\bar{i}k}(x(s),\bar{x}(t_0))\ddot{x}^{\bar{i}} + c_{\bar{i}\bar{j}k}(x(s),\bar{x}(t_0))\dot{x}^{\bar{i}}\dot{x}^{\bar{j}})\dot{x}^k. \end{aligned}$$

When $s = 0$, the last line vanishes by the geodesic equation for $t \in [0,1] \longrightarrow (x(0), \bar{x}(t))$, and the preceding line is manifestly zero. Thus the strictly convex function $\partial^2 g/\partial t^2|_{t=t_0}$ must be nonnegative for $s \in [0,1]$ and, as initially claimed, $\ddot{f}(t_0) = \partial^2 g/\partial t^2|_{(s,t)=(1,t_0)}$ is non-negative at any $t_0 \in]0,1[$ where $\dot{f}(t_0) = 0$.

If the relevant cross-curvature of the cost is positive at one point $(x(s_0), \bar{x}(t_0))$, then the non-negative function $\partial^2 g/\partial t^2|_{t=t_0}$ is strictly convex (4.13) on an interval around $s_0 \in [0,1]$; since $\partial^2 g/\partial t^2|_{t=t_0}$ is minimized at $s = 0$ it must then be positive at $s = 1$, to conclude the proof. □

Definition 4.7. (Illuminated set) *Given $(x, \bar{x}) \in N$, let $V(x, \bar{x}) \subset M$ denote those points $y \in N(\bar{x})$ for which there exists a curve from $(x, \bar{x})$ to $(y, \bar{x})$ in $N(\bar{x}) \times \{\bar{x}\}$ satisfying the geodesic equation on (N, h).*

As a warm up to the *double mountain above sliding mountain* Theorem 4.10, let us derive a strong version of this result under the simplifying hypothesis that the cost is strictly **(A3s)** and not merely weakly regular **(A3w)**.

Corollary 4.8. (Strict maximum principle) *Use a strictly regular cost $c \in C^4(N)$ to define a pseudo-metric (2.1) on the domain $N \subset M \times \bar{M}$. Let $t \in [0,1] \longrightarrow (x, \bar{x}(t)) \in N$ be a geodesic in (N, h) with $\dot{\bar{x}}(0) \neq 0$. Then for all $y \in \cap_{t\in[0,1]} V(x, \bar{x}(t))$ the function $f(t) = -c(y, \bar{x}(t)) + c(x, \bar{x}(t))$ satisfies $f(t) < \max\{f(0), f(1)\}$ on $0 < t < 1$.*

Proof. Let $t \in [0,1] \longrightarrow (x, \bar{x}(t)) \in N$ be a geodesic in (N, h) with $\dot{\bar{x}}(0) \neq 0$. Given $y \in \cap_{t\in[0,1]} V(x, \bar{x}(t))$ and set $f(t) = -c(y, \bar{x}(t)) + c(x, \bar{x}(t))$. Notice $y, x \in N(\bar{x}(t))$ so f is C^4 smooth on $[0,1]$. Proposition 4.6 asserts $\ddot{f}(t_0) > 0$ at each interior critical point $\dot{f}(t_0) = 0$. Any critical point of f in $]0,1[$ is therefore a local minimum, and f is strictly monotone away from this point. Thus $f(t) < \max\{f(0), f(1)\}$ for $0 < t < 1$, as desired. □

If N is horizontally convex then $V(x, \bar{x}(t)) = N(\bar{x}(t))$, which motivates the relation of this corollary to Theorem 3.1. Let us now show a weak version of this maximum principle survives as long as the cost is weakly regular. To handle this relaxation we use a level set approach.

Lemma 4.9. (Level set evolution) *Let $g \in C^2(]\epsilon,\epsilon[\times U)$ where $U \subset \mathbf{R}^n$ is open. Suppose $Dg = (\partial_1 g, \dots, \partial_n g)$ is non-vanishing on $]\epsilon,\epsilon[\times U$. Then the zero set $S(t) = \{x \in U \mid g(t,x) = 0\}$ is a C^2-smooth $n-1$ dimensional submanifold of U which can be parameterized locally for small enough t by $\{X(t,z) \mid z \in S(0)\}$, where the Lagrangian variable $X(t,x)$ solves the ordinary differential equation*

$$\frac{\partial X(t,x)}{\partial t} = -\left[\frac{\partial g}{\partial t}\frac{Dg}{|Dg|^2}\right]_{(t,X(t,x))} \tag{4.14}$$

subject to the initial condition $X(t,x) = x$. Moreover, the positivity set $S^+(t) = \{x \in U \mid g(t,x) \geq 0\}$ has $S(t)$ as its boundary, and expands with an outward normal velocity given by

$$v = -\frac{\partial X}{\partial t}\bigg|_{(t,z)} \cdot \frac{Dg}{|Dg|}\bigg|_{(t,X(t,z))} = \frac{\partial g/\partial t}{|Dg|}\bigg|_{(t,X(t,z))}. \tag{4.15}$$

Proof. Clearly the boundary of $S^+(t)$ is contained in $S(t)$. Since $Dg \neq 0$, the implicit function theorem implies $S(t)$ is a C^2-smooth hypersurface and separates regions where $g(t,x)$ takes opposite signs. Thus $S(t)$ is contained in and hence equal to the boundary in U of the positivity set. If the desired parametrization exists it must satisfy $0 = g(t, X(t,z))$. Differentiation in time yields an equation

$$0 = \frac{\partial g}{\partial t}(t, X(t,z)) + Dg(t, X(t,z)) \cdot \frac{\partial X}{\partial t}\bigg|_{(t,z)}$$

easily seen to be equivalent to (4.14). Conversely, near a point $(0,z) \in]-\epsilon,\epsilon[\times S(0)$, the C^1 vector field (4.14) can be integrated for a short time (depending on z) to yield the desired parametrization. $\square$

Theorem 4.10. (Double mountain above sliding mountain) *Use a weakly regular cost $c \in C^4(N)$ to define a pseudo-metric (2.1) on the domain $N \subset M \times \bar{M}$. Let $t \in]0,1[\longrightarrow (x, \bar{x}(t)) \in N$ be a geodesic in (N,h). If $]t_0,t_1[\times \{y\}$ lies in the interior of*

$$\Lambda := \{(t,y) \in [0,1] \times M \mid y \in V(x, \bar{x}(t))\}, \tag{4.16}$$

then $f(t) = -c(y,\bar{x}(t)) + c(x,\bar{x}(t)) \leq \max\{f(t_0^+), f(t_1^-)\}$ on $0 \leq t_0 < t < t_1 \leq 1$, where

$$f(t_0^+) = \lim_{\epsilon \searrow 0} f(t_0 + \epsilon), \qquad f(t_1^-) = \lim_{\epsilon \searrow 0} f(t - \epsilon). \tag{4.17}$$

Proof. Fix a geodesic $t \in]0,1[\longrightarrow (x, \bar{x}(t)) \in N$ with $\dot{\bar{x}}(1/2) \neq 0$, since otherwise the conclusion is obvious. Note that c-Exp_x and hence $t \in]0,1[\longrightarrow \bar{x}(t)$ are C^3 smooth, from Lemma 4.4 and **(A0)**. For all $y \in M$ and $t \in]0,1[$ set

$$f(t,y) = -c(y,\bar{x}(t)) + c(x,\bar{x}(t))$$

and note that $f(t,y)$ is C^3-smooth on the interior of $\Lambda \subset [0,1] \times M$. Define

$$\begin{aligned} S^+ &:= \{(t,y) \in \operatorname{int}\Lambda \mid \frac{\partial f}{\partial t} \geq 0\} \\ S &:= \{(t,y) \in \operatorname{int}\Lambda \mid \frac{\partial f}{\partial t} = 0\}. \end{aligned}$$

For each $t \in]0,1[$, we think of $f(t,y)$ as defining the elevation of a landscape over M, which evolves from $f(y,0)$ to $f(y,1)$ as t increases, and is normalized so that $f(x,t) = 0$. We picture $f(y,t)$ as a sliding mountain, with $S^+(t) := \{y \in M \mid (t,y) \in S^+\}$ denoting the rising region, and $S(t) := \{y \in M \mid (t,y) \in S\}$ the region at the boundary of $S^+(t)$ which — instantaneously — is neither rising nor sinking.

We claim the rising region $S^+(t) \subset M$ is a non-decreasing function of $t \in]t_0, t_1[$. To see this, we plan to apply Lemma 4.9 to the C^2 function

$$g(t,y) := \frac{\partial f}{\partial t} = -\bar{D}c(y, \bar{x}(t))\dot{\bar{x}}(t) + \bar{D}c(x, \bar{x}(t))\dot{\bar{x}}(t) \tag{4.18}$$

on Λ. Differentiating this function with respect to $y \in M$ yields

$$\begin{aligned} Dg(t,y) &= -D\bar{D}c(y, \bar{x}(t))\dot{\bar{x}}(t) \\ &\neq 0 \end{aligned}$$

which is non-vanishing because of **(A2)**. Applying Lemma 4.9 on any coordinate chart in M shows $S(t)$ is the boundary of $S^+(t)$, and the question of whether $S^+(t)$ is expanding or contracting along its boundary is determined by the sign of $\partial g/\partial t$ on $S(t)$.

From the definition of $V(x, \bar{x}(t'))$ observe $y \in S(t') \subset V(x, \bar{x}(t'))$ implies $(y, \bar{x}(t'))$ is linked to $(x, \bar{x}(t'))$ by a curve in $N(\bar{x}(t')) \times \{\bar{x}(t')\}$ which is geodesic in (N, h). Proposition 4.6 asserts $\partial^2 f/\partial t^2|_{(t',y)} \geq 0$, so from Lemma 4.9 there is a neighbourhood of y on which the rising region $S^+(t)$ does not shrink for a short time interval around $t' \in]0,1[$.

Finally, fix $y \in M$ and $0 \leq t_0 < t_1 \leq 1$ such that $]t_0, t_1[\times \{y\} \subset \operatorname{int} \Lambda$ and let $U \subset]t_0, t_1[$ denote the open set of times at which $y \notin S^+(t)$. If U is non-empty, we claim any connected component of U has $t = t_0$ in its closure. If not, let $t' \in]t_0, t_1[$ denote the left endpoint of a connected component in U. This means $y \in S^+(t')$ but $y \notin S^+(t' + \delta)$ for any $\delta > 0$, in violation of the non-shrinking property of $S^+(t)$ derived above. Thus $t \in]t_0, t_1[\longrightarrow f(t,y)$ is decreasing on an interval $U =]t_0, t(y)[$ for some $t(y) \in [t_0, t_1]$ and non-decreasing on the complementary interval $]t(y), t_1[$. The limits (4.17) exist and the proof that $f(t,y) \leq \max\{f(t_0^+), f(t_1^-)\}$ is complete. $\square$

Remark 4.11 (A geodesic hypersurface bounds the rising region). Lemma 4.4 implies $S(t) \times \{\bar{x}(t)\}$ is a totally geodesic submanifold for each $t \in]0,1[$ of the preceding proof. Indeed, (4.18) shows $\bar{q}^* \in \mathrm{Dom}(c^*\text{-}\mathrm{Exp}_{\bar{x}(t)}) \subset T^*_{\bar{x}(t)}\bar{M}$ belongs to the hyperplane $(\bar{D}c(x, \bar{x}(t)) + \bar{q}^*)\dot{\bar{x}}(t) = 0$, if and only if $y := c^*\text{-}\mathrm{Exp}_{\bar{x}(t)}\bar{q}^*$ lies on $S(t)$.

Remark 4.12. If the domain N is horizontally convex then $V(x, \bar{x}) = N(\bar{x})$ and Λ from (4.16) are open sets. Then $]0,1[\times \{y\} \in \Lambda$ if and only if $y \in \cap_{0<t<1} N(\bar{x}(t))$.

Remark 4.13 (Enhancements). Villani has incorporated a version of our proof into [41], adding simplifications which allow him to avoid the use of the level set method. He further develops our technique to prove additional results: contrast our Theorem A.10, which gives a new approach to a result of Loeper, with Villani's Theorem 12.41, which gives a new approach to a result of Trudinger & Wang.

5. Perspective and conclusions

Before concluding this paper, let us briefly review the connection of the transportation problem (1.1) we study with fully non-linear partial differential equations. Although this connection goes almost back to Monge [30], it has developed dramatically since the work of Brenier [3]. When an optimal mapping $F : M \longrightarrow \bar{M}$ exists and happens to be a diffeomorphism, it provides a change of variables between (M, ρ) and $(\bar{M}, \bar{\rho})$, hence its Jacobian satisfies the equation

$$\bar{\rho}(F(x))|\det DF(x)| = \rho(x). \tag{5.1}$$

Often when F is not smooth, (5.1) remains true almost everywhere [28] [12] [1].

Optimality implies that $F(\cdot)$ can be related to pair of scalar functions $u : M \longrightarrow \mathbf{R} \cup \{+\infty\}$ and $\bar{u} : \bar{M} \longrightarrow \mathbf{R} \cup \{+\infty\}$, which arise from the linear program dual to (1.1), and represent Lagrange multipliers for the prescribed densities ρ and $\bar{\rho}$. Moreover, u can be taken to belong to the class of *c-convex* functions, meaning it can be obtained as the upper envelope of a family of mountains — analogous to (3.1) and (3.8) but with i ranging over $\bar{\rho}$-almost all of $\bar{M}$; similarly, $\bar{u}$ is a c^*-convex function, as in Definition A.3 below. When the twist condition holds, $F = c\text{-Exp} \circ Du$. If non-degeneracy **(A2)** also holds, then (5.1) becomes an equation of Monge-Ampère type expressed as in [27] using local coordinates around $(x', F(\bar{x}'))$ by

$$\det[u_{ij} + c_{ij}]_{(x, c-\mathrm{Exp}_x Du(x))} = \frac{\rho(x)}{\bar{\rho}(c\text{-}\mathrm{Exp}_x Du(x))} |\det c_{i\bar{k}}|_{(x, c-\mathrm{Exp}_x Du(x))}; \tag{5.2}$$

c-convexity of u implies non-negative definiteness of the matrix $u_{ij} + c_{ij}$ of second derivatives, hence the equation is degenerate elliptic [19]. Without further assumptions on (M, ρ) and the geometry of $(\bar{M}, \bar{\rho})$ we can expect neither strict ellipticity nor regularity of solutions: if the support of ρ is connected but the support of $\bar{\rho}$ is not, this will force F to be discontinuous and $u \notin C^1(M)$. The equation is therefore not locally smoothing, and the best one can hope is for solutions to inherit regularity from the boundary data (M, ρ) and $(\bar{M}, \bar{\rho})$.

For the cost function $c(x, \bar{x}) = |x - \bar{x}|^2/2$ on $M, \bar{M} \subset \mathbf{R}^n$, (5.2) becomes the familiar Monge-Ampère equation for the convex function $u(x) + |x|^2/2$ [3]. In this case Caffarelli [5] was able to show *convexity* of $\bar{M}$ implies local Hölder continuity of F if the densities $d\rho/d\,\mathrm{vol} \in L^\infty(M)$ and $d\,\mathrm{vol}/d\bar{\rho} \in L^\infty(\bar{M})$ are bounded, and smoothness of F on the interior of M if the densities $\log|d\rho/d\,\mathrm{vol}\,|$ and $\log|d\bar{\rho}/d\,\mathrm{vol}\,|$ are bounded and smooth; see also Delanoë [14] for the case $n = 2$, and Wang [42] for analogous results in the context of Example 3.5, the reflector antenna problem. The results of Loeper and Ma, Trudinger & Wang extend the Hölder [26] and smooth [27] [35] theories to general bi-twisted, strictly regular costs on $\mathrm{cl}(M \times \bar{M})$. Loeper in particular achieves stronger results such as a global Hölder estimate with explicit exponent under weaker restrictions on ρ and $\bar{\rho}$ by exploiting strict regularity of the cost; see Theorems D.1 and E.3. Trudinger & Wang extended the up-to-the-boundary regularity results of Caffarelli [4] and Urbas [36] — which require convexity and smoothness of $M \subset \mathbf{R}^n$ as well as $\bar{M}$ — to bi-twisted costs which are merely weakly regular [34]. For bounded and sufficiently smooth densities,

horizontal and vertical convexity of $N = M \times \bar{M}$ takes the place of the convexity assumptions on the target $\bar{M}$ and domain M in these theories.

Since we had not previously encountered pseudo-Riemannian geometry of any signature other than the Lorentzian one $(n, 1)$ in applications, much less as the necessary and sufficient condition for degenerate elliptic partial differential equations to possess smooth solutions, a few words of explanation seem appropriate.[2]

The regularity of optimal maps $F : M \longrightarrow \bar{M}$ is a question whose answer should depend only on the cost function $c(x, \bar{x})$ and the probability measures ρ and $\bar{\rho}$. It should not depend on which choice of smooth coordinates on M and $\bar{M}$ are used to represent this data or the solution. Any necessary and sufficient condition on $c(x, \bar{x})$ guaranteeing regularity of F should therefore be geometrically invariant, meaning coordinate independent. Thus pseudo-Riemannian geometry and curvatures arise in the theory of optimal transportation for the same reason they arise in Einstein's theory of gravity, *general relativity*: they provide the natural language for describing phenomena — in this case regularity — which exhibit invariance under the general group of diffeomorphisms. Put another way, the underlying physical reality is independent of whose coordinates are used to describe it. Moreover, this invariance places severe restrictions on the form which necessary and sufficient conditions for regularity can take: since the pseudo-metric h is equivalent to knowing the cost function $c(x, \bar{x})$ — up to null Lagrangians $v(x) + \bar{v}(\bar{x})$ which are irrelevant to the optimization at hand (1.1) — any such conditions on the cost function must be expressible via the curvature tensor of h.

Let us turn now to the question of why the intrinsic geometry of optimal transportation should be pseudo-Riemannian rather than Riemannian. Dimensional symmetry between the domain M and target $\bar{M}$ suggest that the number of time-like directions in the theory — if any — should equal the number of space-like directions. But why signature (n, n) rather than the signature $(2n, 0)$, which is more frequently associated to elliptic and extremal problems? And why the nullity of $p \oplus \bar{p}$ in **(A3w)**?

The Riemannian notion of length allows us to associate a magnitude to any sectional curvature. However, null vectors have no length and cannot be normalized; because the plane $(p \oplus 0) \wedge (0 \oplus \bar{p})$ is generated by orthogonal null-vectors, we can decide the sign (positive, negative, or zero) of its sectional curvature, but not the magnitude. On the other hand, the results of Loeper reveal that the size of the constant $C > 0$ in hypothesis (1.2) controls the Hölder constant of the mapping $F : M \longrightarrow \bar{M}$. Unlike the exponent, which is coordinate-independent, the Hölder constant of F obviously depends on the choice of coordinates. We are therefore relieved to find the cross-curvature condition governing regularity does not have an

[2] However, as we learned subsequently from Robert Bryant, the wedge product $\omega \wedge \omega$ can be used to define a signature $(3, 3)$ pseudo-metric on the space $\mathbf{R}^4 \wedge \mathbf{R}^4$; in four dimensions, the difference between *positive sectional curvature* and *positive curvature operator* amounts to the question of whether the curvature operator is positive-definite only on the light cone with respect to this pseudo-metric, or on the full space; c.f. [2]. In one way this parallels the distinction between strict regularity and positive cross-curvature of a cost; in another it parallels the distinction between positive cross-curvature and positive sectional curvature, (2.3)– (2.4).

associated magnitude, since the problem has no intrinsic length scale. To be scale free, the geometrical structure which governs regularity for optimal transportation must be pseudo-Riemannian, since the modulus of continuity of a map F has no intrinsic meaning in the absence of separate notions of length on M and $\bar{M}$, which the cost function $c(x,\bar{x})$ alone cannot provide. What it can and does provide are geodesics on $N(\bar{x}) \times \{\bar{x}\}$ and $\{x\} \times N(x)$, and geodesic convexity of these null submanifolds is the essential domain hypothesis in Ma, Trudinger & Wang's theory.

In the present investigation, we have focused exclusively on the pseudo-metric $d\ell^2 = -c_{i\bar{j}} dx^i dx^{\bar{j}}$ induced by the cost (2.1). Let us conclude by noting that there is also a canonical symplectic form $\omega = d(Dc \oplus 0) = -d(0 \oplus \bar{D}c)$ on $N \subset M \times \bar{M}$ associated to the cost $c \in C^4(N)$. In local coordinates $x^1, \ldots, x^n$ on M and $x^{\bar{1}}, \ldots, x^{\bar{n}}$ on $\bar{M}$ it is given by

$$\omega := \frac{1}{2} \begin{pmatrix} 0 & \bar{D}Dc \\ -D\bar{D}c & 0 \end{pmatrix}. \tag{5.3}$$

It is possible to verify that any c-optimal diffeomorphism $F : M \longrightarrow \bar{M}$ has a graph which is spacelike with respect to h and Lagrangian with respect to ω. Conversely, for a weakly regular cost, results of Trudinger & Wang [34] or Villani [41] can be used to deduce that any diffeomorphism whose graph is h-spacelike and ω-Lagrangian is in fact the c-optimal map between the measures $\rho := \pi_\#(\mathrm{vol}^{(N,h)}|_{\mathrm{Graph}(F)})$ and $\bar{\rho} := \bar{\pi}_\#(\mathrm{vol}^{(N,h)}|_{\mathrm{Graph}(F)})$ obtained by projecting the Riemannian volume $\mathrm{vol}^{(N,h)}$ induced by h on $\mathrm{Graph}(F)$ through the canonical projections $\pi(x,\bar{x}) = x$ and $\bar{\pi}(x,\bar{x}) = \bar{x}$. This reveals another unexpected connection between optimal transportation and symplectic (or pseudo-Kähler) geometry. When ρ and $\bar{\rho}$ are given by the Euclidean volumes on two convex domains, and $c(x,\bar{x}) = |x-\bar{x}|^2/2$, this is related to the work of Wolfson [45] and Warren [44] on special Lagrangian submanifolds, where a pseudo-Riemannian metric of signature (n,n) also appears [44]. We defer the details of this development to a future work.

Appendix A. Regularity consequences for optimal transportation

This series of appendices is devoted to explaining how the results obtained above combine with Loeper's ideas to simplify the proof of his Hölder continuity results for optimal mappings with respect to cost functions which are strictly regular **(A3s)** — meaning they have positive cost-sectional curvature in the language of Loeper [26]. To emphasize that this argument is completely self-contained, we recall the necessary details of his proof in full, exploiting where possible the conceptual framework developed above. We often present variations on his arguments, but do not claim originality for the conclusions.

The plan can be outlined as follows. In the present section we recall the relevant facts of life from the literature, and derive the key local-implies-global ingredient — a variant of which appears in Trudinger & Wang [35] — as a simple corollary to our main theorem above. In Appendix B we recall Loeper's Taylor expansion, which uses strict positivity of orthogonal cross-curvatures to quantify the altitude of the double mountain over the sliding mountain near the point in the valley of

indifference where the sliding mountains are normalized to coincide. Appendices C and D explain how Loeper used this result to obtain Hölder continuity of optimal mappings between suitable measures on Euclidean domains. Appendix E employs the same argument in a simpler setting to obtain continuity of optimal mappings between suitable measures on the round sphere, when optimality is measured either against (a) Riemannian distance squared [26], or (b) the reflector antenna cost function, as in [26] and Caffarelli, Gutiérrez & Huang [8]. As a corollary, we recover directly a bound on how closely an optimal map may approach the cut locus, which can be used in place of Delanoë & Loeper's estimate [17] in the next step. Once the continuity of the optimal map has been established, Caffarelli, Gutiérrez, Huang and Loeper's Hölder regularity results concerning optimal maps on the spheres (a) [26] and (b) [8] [26] can be recovered using bi-twisting **(A1)** alone without further resort to the local-implies-global theorem: the problem is easily localized when continuity is known. This approach is considerably more direct than Loeper's original argument, which established the local-implies-global theorem using approximation by smooth solutions to a family of auxiliary problems constructed by combining Delanoë's continuity method [15] with Delanoë & Loeper's control on proximity to the cut locus [17], and the a priori estimates of Ma, Trudinger & Wang [27].

A.1. ***c*-convex functions and their properties.** The definition of c-convex functions and some basic properties are discussed. We first introduce the notion of a supporting mountain (or c-mountain), also called c-support functions, c-planes or c-graphs. The focus (or c-focus) of such a mountain is defined below, and was called the c-normal by Trudinger & Wang, in analogy with the Gauss map and generalized normal in the theory of convex bodies.

We shall find it convenient to continue to assume M and $\bar{M}$ are open manifolds whose closures $\mathrm{cl}\, M$ and $\mathrm{cl}\, \bar{M}$ are compact in some larger space, while $N \subset M \times \bar{M}$ intersects neither $M \times \partial \bar{M}$ nor $(\partial M) \times \bar{M}$.

Definition A.1. (Supporting mountains) *Fix* $c : \mathrm{cl}(M \times \bar{M}) \longrightarrow \mathbf{R} \cup \{+\infty\}$. *A* mountain *refers to any function* f *on* $\mathrm{cl}\, M$ *of the form*

$$f(\cdot) = -c(\cdot, \bar{x}) + \lambda$$

with $(\lambda, \bar{x}) \in \mathbf{R} \times \mathrm{cl}\, \bar{M}$. *The mountain is said to be* focused *at* $\bar{x}$. *The mountain* f *is said to* support $u : \mathrm{cl}\, M \longrightarrow \mathbf{R} \cup \{+\infty\}$ *at* $x \in \mathrm{cl}\, M$ *if* $u(x) = f(x) < +\infty$ *and*

$$u(y) \geq f(y)$$

for all $y \in \mathrm{cl}\, M$. *The mountain is said to support* u to first order *at* $x \in M$ *if* $u(x) = f(x) < +\infty$, *and* M *has a manifold structure near* x *with*

$$u(y) \geq f(y) + o(|y - x|) \qquad \text{as } y \to x. \tag{A.1}$$

Remark A.2 (Twisting identifies focus). If a mountain f supports u to first order at a point $x \in M$ where u happens to be differentiable, and f is known to be focused in $\bar{N}(x)$, then twisting **(A1)** identifies the focus $\bar{x} = c\text{-}\mathrm{Exp}_x Du(x)$ uniquely, since (A.1) then implies $Du(x) = Df(x) = -Dc(x, \bar{x})$.

Definition A.3. (c-convex function focused in $\bar\Omega$) *Fix $c : \mathrm{cl}(M \times \bar M) \longrightarrow \mathbf{R} \cup \{+\infty\}$ and a subset $\bar\Omega \subset \mathrm{cl}\,\bar M$. Then $u : \mathrm{cl}\,M \longrightarrow \mathbf{R} \cup \{+\infty\}$ belongs to $\mathcal{S}^{-c}_{\bar\Omega}$ if $\mathrm{Dom}\,u := \{x \in M \mid u(x) < +\infty\}$ is non-empty and there exists $v : \bar\Omega \longrightarrow \mathbf{R} \cup \{+\infty\}$ such that $u = v^c_{\bar\Omega}$, where*

$$v^c_{\bar\Omega}(x) := \sup_{\bar x \in \bar\Omega} -c(x,\bar x) - v(\bar x).$$

Functions in the class $\mathcal{S}^{-c}_{\bar\Omega}$ are sometimes called *c-convex*, though they might also be called *$(-c)$-convex*, since they are *supremal convolutions* with $-c$. Our subscript keeps track of the set $\bar\Omega$ of foci; we say $u \in \mathcal{S}^{-c}_{\bar\Omega}$ is a c-convex function focused in $\bar\Omega$.

Lemma A.4. (Properties c-convex functions inherit from the cost) *Fix metric spaces (M,d) and $(\bar M,\bar d)$, a cost $c : \mathrm{cl}(M \times \bar M) \longrightarrow \mathbf{R} \cup \{+\infty\}$ and subset $\bar\Omega \subset \mathrm{cl}\,\bar M$. Recall $c^*(\bar x, x) := c(x,\bar x)$.*
(a) Legendre-type duality: *If $u \in \mathcal{S}^{-c}_{\bar\Omega}$ then $u = (u^{c^*}_{\mathrm{cl}\,M})^c_{\bar\Omega}$.*
(b) Semicontinuity: *If $c : \mathrm{cl}\,M \times \bar\Omega \longrightarrow \mathbf{R}$ is continuous and $u \in \mathcal{S}^{-c}_{\bar\Omega}$, then $u : \mathrm{cl}\,M \longrightarrow \mathbf{R} \cup \{+\infty\}$ is lower semicontinuous and $\partial^c u \subset \mathrm{cl}(M \times \bar M)$ is closed.*
(c) Lipschitz: *If $\Omega \subset \mathrm{cl}\,M$ and $u \in \mathcal{S}^{-c}_{\bar\Omega}$ then*

$$\mathrm{Lip}(u,\Omega) := \sup_{\Omega \ni y \neq z \in \Omega} \frac{u(y)-u(z)}{d(y,z)} \le \sup_{\bar x \in \bar\Omega} \mathrm{Lip}(-c(\cdot,\bar x),\Omega).$$

(d) Semiconvex: *If $(M,d) = (\mathbf{R}^n, |\cdot|)$ and $D^2 c(\cdot,\bar x) \le CI$ holds in the (matrix) sense of distributions on a domain $\Omega \subset \mathbf{R}^n$, with $C = C(\Omega,\bar\Omega) \in \mathbf{R}$ independent of $\bar x \in \bar\Omega$, then each $u \in \mathcal{S}^{-c}_{\bar\Omega}$ satisfies $D^2 u \ge -CI$ in the sense of distributions on Ω.*

Proof. All four facts are well-known; (b) is elementary to check and the other three statements are proved, e.g., in [41], where (a) is Proposition 5.8, and (c) and (d) are established in the proof of Theorem 10.26. ☐

Corollary A.5. (Lipschitz/semiconvex) *If c is a locally Lipschitz function on the (compact) closure of $\mathrm{cl}(M \times \bar M)$, then $u \in \mathcal{S}^{-c}_{\mathrm{cl}\,\bar M}$ is locally Lipschitz on $\mathrm{cl}\,M$ and $\mathrm{cl}\,M \subset \mathrm{Dom}\,\partial^c u$. If, in addition, c is locally semiconcave, then $u \in \mathcal{S}^{-c}_{\mathrm{cl}\,\bar M}$ is locally semiconvex (A.2) on M.*

Proof. The fact that $u \in \mathcal{S}^{-c}_{\mathrm{cl}\,\bar M}$ inherits Lipschitz or semiconvexity properties from the cost follows directly from Lemma A.4(c) and (d), using compactness of $\mathrm{cl}\,\bar M$ to conclude that the local semiconcavity constant C of $c(\cdot,\bar x)$ and Lipschitz constant $\mathrm{Lip}(-c(\cdot,\bar x),\Omega)$ are independent of $\bar x \in \mathrm{cl}\,\bar M$. To see $\mathrm{cl}\,M \subset \mathrm{Dom}\,\partial^c u$, use Lemma A.4(a) to write

$$u(x) = \sup_{\bar x \in \mathrm{cl}\,\bar M} -c(x,\bar x) - u^{c^*}_{\mathrm{cl}\,M}(\bar x).$$

Here $u^{c^*}_{\mathrm{cl}\,M}$ is lower semicontinuous by Lemma A.4(b) so the supremum is attained at some $\bar x \in \mathrm{cl}\,\bar M$, whence

$$\begin{aligned} u(x) + c(x,\bar x) &= -u^{c^*}_{\mathrm{cl}\,M}(\bar x) \\ &= -\sup_{y \in \mathrm{cl}\,M} -c(y,\bar x) - u(y) \\ &\le u(y) + c(y,\bar x) \end{aligned}$$

for all $y \in \operatorname{cl} M$. Thus $(x, \bar{x}) \in \partial^c u$ as desired (2.8). $\square$

Note that c is both locally Lipschitz and semiconcave if it is weakly regular on $\operatorname{cl}(N)$, or else if $c = d^2/2$ is the distance squared on a compact Riemannian manifold $(M = \bar{M}, g)$ without boundary as in Example 3.6; see Cordero-Erausquin, McCann, Schmuckenschläger [13]. This is not the case for the reflector antenna cost function of Example 3.5, but the associated c-convex functions are Lipschitz and semiconvex anyways — unless they are mountains — as the following proposition shows. Here, as above, semiconvexity of $u : M \longrightarrow \mathbf{R}$ means that near each point $z \in M$ there is a coordinate ball $B_r(z) \subset M$ and coordinate-dependent constant $C < \infty$ such that $D^2 u(x) \geq -C\mathtt{I}$ as matrices distributionally in these coordinates, or equivalently that

$$x \in B_r(z) \longrightarrow u(x) + C|x|^2/2 \tag{A.2}$$

is a convex function, where $|x|$ denotes the Euclidean norm of the coordinates. Semiconcavity of u means $-u$ is semiconvex.

Proposition A.6. (Semiconvexity of reflector antenna potentials) *Fix the cost $c(x, \bar{x}) = -\log|x - \bar{x}|$ on the Euclidean unit sphere $\mathbf{S}^n = \partial B_1(0) \subset \mathbf{R}^{n+1}$. If $u : \mathbf{S}^n \longrightarrow \mathbf{R} \cup \{+\infty\}$ belongs to $\mathcal{S}^{-c}_{\mathbf{S}^n}$ and takes finite values at two or more points, then u is locally Lipschitz and semiconvex on $\mathbf{S}^n$. In coordinates, its Lipschitz and semiconvexity constants do not depend on u except through* $\max_{\mathbf{S}^n} u - \min_{\mathbf{S}^n} u$.

Proof. Since $u \in \mathcal{S}^{-c}_{\mathbf{S}^n}$, defining the dual potential $\bar{u}(\cdot) = \sup_{x \in \mathbf{S}^n} \log|\cdot - x| - u(x)$ yields

$$u(\cdot) = \sup_{\bar{x} \in \mathbf{S}^n} \log|\cdot - \bar{x}| - \bar{u}(\bar{x}) \tag{A.3}$$

similarly to Lemma A.4(a). We shall show that both u and $\bar{u}$ are bounded from below and above. Since $u(\cdot)$ takes finite values at $x_0 \neq x_1$ in $\mathbf{S}^n$, we have $\bar{u}(\cdot) \geq \max_{i=0,1} \log|\cdot - x_i| - u(x_i) \geq \mathit{const} > -\infty$, where the constant depends only on $\max\{u(x_0), u(x_1)\}$ and $|x_0 - x_1|$. Thus $\bar{u}(\cdot)$ is bounded below, hence $u(\cdot)$ is bounded above from (A.3) since $|x - \bar{x}| \leq 2$ for all $x, \bar{x} \in \mathbf{S}^n$.

On the other hand, if $\operatorname{Dom} \bar{u} := \{\bar{x} \in \mathbf{S}^n \mid \bar{u}(\bar{x}) < \infty\}$ consists of one or fewer points $\bar{y}$, we reach the contradiction $u(\bar{y}) = -\infty$. So $\operatorname{Dom} \bar{u}$ consists of two or more points, and the same argument as before yields $u(\cdot)$ bounded below and $\bar{u}(\cdot)$ bounded above.

To show u is semiconvex, recall that any supremum of smooth functions which have locally uniform control on their C^2 norms is semiconvex, as in Lemma A.4(d). Let $-\bar{x}$ denote the antipodal point to a fixed point $\bar{x} \in \mathbf{S}^n$. If $|\lambda_0 - \lambda_1| \leq \Lambda$, the function

$$v(\cdot) := \max\{\log|\cdot - \bar{x}| - \lambda_0, \log|\cdot + \bar{x}| - \lambda_1\}$$

is semiconvex on $\mathbf{S}^n$, with a semiconvexity constant C depending only on $\Lambda < +\infty$, but independent of $\lambda_0, \lambda_1 \in \mathbf{R}$ and $\bar{x} \in \mathbf{S}^n$. Setting $\Lambda = (\sup_{\mathbf{S}^n} u) - (\inf_{\mathbf{S}^n} u)$, we see from (A.3) that

$$u(\cdot) = \sup_{\bar{x} \in \mathbf{S}^n} \max\{\log|\cdot - \bar{x}| - \bar{u}(\bar{x}), \log|\cdot + \bar{x}| - \bar{u}(-\bar{x})\}$$

is a supremum of such functions $v(\cdot)$, hence semiconvex with constant C as desired. Finally, any bounded convex function is locally Lipschitz, hence bounded semiconvex functions are locally Lipschitz by (A.2). $\square$

Definition A.7. (Super- and subdifferential) *The* subdifferential $\partial u \subset T^*M$ *of a function* $u : M \longrightarrow \mathbf{R} \cup \{+\infty\}$ *on a smooth manifold is defined to consist of those points* $(x, q^*) \in T^*M$ *in the cotangent bundle such that in local coordinates around* x*,*

$$u(y) \geq u(x) + q^*(y - x) + o(|y - x|) \quad \text{as } y \to x. \tag{A.4}$$

It follows that $\partial u(x) := \{q^* \in T^*_x M \mid (x, q^*) \in \partial u\}$ is a convex set, and $\partial u(x) = \{Du(x)\}$ at those points $x \in \operatorname{Dom} Du \subset M \setminus \partial M$ where u is differentiable. The superdifferential $-\partial(-u)$ consists of those pairs $(x, q^*) \in T^*M$ which satisfy the opposite inequality. We set $\operatorname{Dom} \partial u := \{x \in M \mid \partial u(x) \neq \emptyset\}$, so that $\operatorname{Dom} Du = (\operatorname{Dom} \partial u) \cap (\operatorname{Dom} \partial(-u))$. If a function u on M is locally Lipschitz and semiconvex, then $\partial u \subset T^*M$ is closed and $M \subset \operatorname{Dom} \partial u$.

Let us now hypothesize an extension of the twist condition to mountains focused in $\mathrm{cl}(\bar{M})$ as follows.

Definition A.8. (Extended twist) *A cost* $c : \mathrm{cl}(M \times \bar{M}) \longrightarrow \mathbf{R} \cup \{+\infty\}$ *satisfies the extended twist condition* **(A1′)** *if each mountain* f *is subdifferentiable on* $M \cap \operatorname{Dom} f$*, where* $\operatorname{Dom} f := \{y \in \mathrm{cl}\, M \mid f(y) < \infty\}$*, and if there is a continuous map* $c\text{-}Exp' : T^*M \cap \mathrm{cl}(\operatorname{Dom} c\text{-}Exp) \longrightarrow \mathrm{cl}\, \bar{M}$*, such that each* $x \in M$*, mountain* $f(\cdot) = -c(\cdot, \bar{x}) + \lambda$*, and extreme point* p^* *of* $\partial f(x)$ *satisfy* $(x, p^*) \in \mathrm{cl}(\operatorname{Dom} c\text{-}Exp)$ *and* $\bar{x} = c\text{-}Exp'_x p^*$*.*

One way to satisfy **(A1′)** is to assume **(A1)** extends to some larger manifold $\tilde{N}$ in which $M \times \bar{M} \subset\subset \tilde{N}$ is compactly embedded, which we abbreviate by saying **(A1)** holds on $\mathrm{cl}\, N$. Property **(A1′)** is also verified for the examples discussed hereafter: any compact Riemannian manifold $(M = \bar{M}, g)$ with the geodesic distance squared cost $c = d^2/2$ of Example 3.6; and the reflector antenna cost $c(x, \bar{x}) = -\log |x - \bar{x}|$ on the sphere $\mathbf{S}^n \times \mathbf{S}^n$ as in Example 3.5. The key to verification in the Riemannian case is that the Riemannian exponential is continuously defined throughout the tangent bundle, the cut-locus has no interior, and mountains are semiconvex [13], so extreme points in $\partial f(x)$ of a mountain f are limits of gradients $p^* = \lim_{k \to \infty} Df(x_k)$ from nearby points of differentiability $x_k \in \operatorname{Dom} Df$; c.f. §25.6 [32].

The extended twist condition **(A1′)** yields the following lemma and theorem.

Lemma A.9. (Tangent mountains support c-convex functions globally) *Suppose a cost* $c : \mathrm{cl}(M \times \bar{M}) \longrightarrow \mathbf{R} \cup \{+\infty\}$ *twisted on* $N \subset M \times \bar{M}$ *satisfies the extended twist hypothesis* **(A1′)***. If* $u \in \mathcal{S}^{-c}_{\mathrm{cl}\, \bar{M}}$ *is differentiable at* $x \in M \cap \operatorname{Dom} \partial^c u$ *then* $(x, Du(x)) \in \mathrm{cl} \operatorname{Dom} c\text{-}Exp$ *and* $\partial^c u(x) = \{\bar{x}\}$ *with* $\bar{x} := c\text{-}Exp'_x Du(x)$*. Any mountain supporting* u *to first order at* x *is focused at* $\bar{x}$ *and supports* u *globally.*

Proof. Fix a point $x \in M \cap \operatorname{Dom} \partial^c u$ where $u \in \mathcal{S}^{-c}_{\mathrm{cl}\, \bar{M}}$ is differentiable. Then $u(x) < +\infty$ and there exists $\bar{x} \in \partial^c u(x)$ such that the mountain $f(\cdot) = -c(\cdot, \bar{x}) + \lambda$ with $\lambda = c(x, \bar{x}) + u(x)$ supports u at x. Now let $g(\cdot) = -c(\cdot, \bar{y}) + \mu$ be any other

mountain which supports u to first order at x. Then $g(y) \leq u(y) + o(|y-x|) = u(x)+Du(x)(y-x)+o(|y-x|)$ is superdifferentiable at x; since g is subdifferentiable by hypothesis, we conclude $Dg(x) = Du(x)$. In this case $\partial g(x) = \{Du(x)\}$, has $Du(x)$ as an extreme point, so the extended twist hypothesis implies g is focused at $\bar{y} = c\text{-Exp}'_x Du(x)$. Applying the same argument to f instead of g yields $\bar{x} = c\text{-Exp}'_x Du(x)$, hence $\bar{x} = \bar{y}$ and $g = f + const$. The constant vanishes since $g(x) = f(x)$, so we conclude any mountain g which supports u to first order at $x \in \operatorname{Dom} Du$ actually supports u globally. ☐

We are now in a position to extend this lemma to points where u is not differentiable by applying Theorem 3.1. This extension is the key local-implies-global ingredient required for Loeper's argument. Our proof of Theorem A.10 is similar to the proof of Proposition 2.12 [26]. Under the stronger hypothesis **(A3s)**, Trudinger & Wang's corrigendum contains a different approach to a similar result [35] (but with strictly regular costs and a stronger hypothesis on the domains).

Theorem A.10. (Mountains supporting to first-order support globally) *Use a continuous cost $c : \operatorname{cl}(M \times \bar{M}) \longrightarrow \mathbf{R} \cup \{+\infty\}$ which has a twisted and weakly regular* **(A0)–(A3w)** *restriction $c \in C^4(N)$ to define a pseudo-metric (2.1) on a horizontally convex domain $N \subset M \times \bar{M}$. If c is unbounded assume that whenever $(x_k, \bar{x}_k) \in \operatorname{Dom} c$ converges and $c(x_k, \bar{x}_k) \to +\infty$ then*

$$\lim_{k\to\infty} |Dc(x_k, \bar{x}_k)| = +\infty. \tag{A.5}$$

Fix $x \in M$ such that $\operatorname{cl}\bar{M}$ appears convex from x, and suppose $\cap_{0\leq t\leq 1} N(\bar{x}(t))$ is dense in M for each geodesic $t \in [0,1] \longrightarrow (x, \bar{x}(t)) \in \{x\} \times \operatorname{cl}\bar{M}$. Fix $u \in \mathcal{S}^{-c}_{\operatorname{cl}\bar{M}}$ for which there is a coordinate neighbourhood $U \subset M$ of x contained in $\operatorname{Dom}\partial^c u$, and a constant $C < \infty$ such that each mountain $f : \operatorname{cl} M \longrightarrow \mathbf{R}$ which supports u to first order at some $y \in U$ satisfies $|\nabla f| \leq C$ and $D^2 f \geq -C\mathtt{I}$ distributionally in the given coordinates on U. If the extended twist hypothesis **(A1′)** *holds, any mountain which supports u to first order at x will be dominated by u throughout $\operatorname{cl} M$.*

Proof. Take $u \in \mathcal{S}^{-c}_{\operatorname{cl}\bar{M}}$ and $x \in U \subset \operatorname{Dom}\partial^c u \setminus \partial M$ as in the theorem. Assume $x \notin \operatorname{Dom} Du$, since otherwise the conclusion follows immediately from Lemma A.9. We see u is Lipschitz and semiconvex on U by Lemma A.4, after noting that u agrees throughout U with the supremum of those mountains $f \leq u$ which support u on M and make contact with u on $U \subset \operatorname{Dom}\partial^c u$. The subdifferential $\partial u(x) \subset T^*_x M$ is a non-empty convex set from the semiconvexity of u; it is bounded since u is Lipschitz. Let us first show that if p^* is an extreme point of $\partial u(x)$ then $c\text{-Exp}'_x p^* \in \partial^c u(x)$. Extremality implies x is the limit of a sequence $x_k \in U \cap \operatorname{Dom} Du$ such that $p^* = \lim_{k\to\infty} Du(x_k)$, according to §25.6 [32]. Since $\partial^c u(x_k)$ is non-empty, Lemma A.9 yields a sequence of mountains $f_k(\cdot) = -c(\cdot, \bar{x}_k) + c(x_k, \bar{x}_k) + u(x_k)$ supporting u and focused at points $\bar{x}_k = c\text{-Exp}'_{x_k} Du(x_k)$ which converge to a limit $\bar{x} := c\text{-Exp}'_x p^*$. Since $|Dc(x_k, \bar{x}_k)|$ is bounded by the Lipschitz constant C of u on U, (A.5) implies $c(x_k, \bar{x}_k)$ has a finite limit. It then follows from $u(\cdot) \geq f_k(\cdot)$ that $f(\cdot) = -c(\cdot, \bar{x}) + c(x, \bar{x}) + u(x)$ supports u on $\operatorname{cl} M$.

According to Theorem 3.1, the set $\partial^c u(x)$ of foci corresponding to mountains $h(\cdot) = -c(\cdot, \bar{z}) + c(x, \bar{z}) + u(x)$ which support u at x appears convex from x. Thus there is a geodesic in N connecting $(x, \bar{x})$ to $(x, \bar{z})$. From the twist condition, this means set of cotangent vectors q^* such that $c\text{-Exp}'_x q^* \in \partial^c u(x)$ form a convex set in $T^*_x M$. This convex set contains $\partial u(x)$, since it contains the extreme points of $\partial u(x)$. Finally, suppose a mountain $g(\cdot) = -c(\cdot, \bar{y}) + \mu$ supports u to first order at x. The semiconvexity of $g(\cdot)$ yields some q^* extremal in $\partial g(x)$ with $\bar{y} = c\text{-Exp}'_x q^*$. Moreover, $q^* \in \partial u(x)$ a fortiori (A.1), hence there is a mountain f also focused at $\bar{y}$ which supports u at x. Now $g = f$ since their foci and values at x coincide, so the theorem is established. □

Remark A.11 (the sphere, the reflector antenna). Although the hypotheses of Theorem A.10 appear complicated, they are obviously satisfied on any product manifold $N = M \times \bar{M}$ on which **(A0)-(A3w)** extend to a larger manifold $\tilde{N} \supset\supset N$ which contains N compactly. For the distance squared cost $c = d^2/2$ on a compact and boundaryless Riemannian manifold $(M = \bar{M}, g)$, the cost function is globally Lipschitz and semiconvex [13], hence all mountains $f(\cdot) = -d^2(\cdot, \bar{y})/2 + \mu$ obey the desired local estimates. Taking $N \subset M \times \bar{M}$ to be the complement of the cut locus, the only delicacies are bi-convexity of N and the density of $\cap_{0 \le t \le 1} N(\bar{x}(t))$; on the round sphere these two properties are satisfied, as Loeper knew and we discuss elsewhere. On $\mathbf{S}^n \times \mathbf{S}^n$, with the reflector antenna cost $c(x, \bar{x}) = -\log |x - \bar{x}|$ of Example 3.5, the mountains are smooth away from their foci, so choosing the diameter of U small relative to the Lipschitz constant C of u, (A.5) combines with Proposition A.6 to force all mountains which support u to first order in U to have foci well outside of U.

Remark A.12 (products $\mathbf{S}^{n_1} \times \cdots \times \mathbf{S}^{n_k} \times \mathbf{R}^l$ and their Riemannian submersions). Theorem A.10 holds for the distance squared cost on the Riemannian product $M = \bar{M}$ of round spheres $\mathbf{S}^{n_1} \times \cdots \times \mathbf{S}^{n_k} \times \mathbf{R}^l$ or its Riemannian submersions. The necessary hypotheses are satisfied according to Example 3.9 and Remark 3.10.

Appendix B. Estimating the height of the double mountain relative to the sliding mountain near their point of coincidence

The following local estimate is the main result of this section; due to Loeper [26], it is crucial to his argument for Hölder continuity of optimal mappings. One can view it as a localized version of the double mountain above sliding mountain principle (Theorem 4.10), in which strict regularity of the cost is used to quantify the relative altitude of the double mountain over the sliding mountain in a neighbourhood of the point where they coincide. For this entire section, we shall work in a single Euclidean coordinate patch; the norm $|p|^2 = p \cdot p$ and derivatives such as $D^2 c = \{c_{ij}\}$ and $D\bar{D}c = \{c_{i\bar{j}}\}$ refer to these coordinates.

Proposition B.1. (Local double-above-sliding-mountain estimate [26]**)** *Let $c \in C^4(\mathrm{cl}\, N)$ be strictly regular* **(A2)**–**(A3s)** *on the closure of product $N = M \times \bar{M}$*

of two bounded domains M, $\bar M \subset \mathbf{R}^n$, so that the pseudo-metric (2.1) obeys

$$\sec_{(x,\bar x)}(p \oplus 0) \wedge (0 \oplus \bar p) \geq 2C_0 |p|^2 |\bar p|^2 \tag{B.1}$$

for each $(x, \bar x) \in N$ and null vector $p \oplus \bar p \in T_{(x,\bar x)}N$, the norms $|p|$ and $|\bar p|$ being defined using the Euclidean inner product on the global coordinates. Then there exist positive constants $r_0 = r_0(\operatorname{Diam} \bar M, \|[D\bar D c]^{-1}\|_{C^0(N)}, \|c\|_{C^4(N)}, C_0)$ and $C_1 = C_0 \|2D\bar D c\|^{-2}_{C^0(N)} \|[D\bar D c]^{-1}\|^{-2}_{C^0(N)}$ such that: for each coordinate ball $B_r(x) \subset M$ of radius $0 < r < r_0$ and geodesic $t \in [0,1] \longrightarrow (x, \bar x(t)) \in N$ with $\dot{\bar x}(0) \neq 0$, the function $f(t,y) := -c(y, \bar x(t)) + c(x, \bar x(t))$ satisfies

$$\max[f(0,y), f(1,y)] \geq f(t,y) + C_1 t(1-t)|\bar x(1) - \bar x(0)|^2 |y-x|^2 - \|c\|_{C^3(N)} |y-x|^3$$

throughout $(t,y) \in [0,1] \times B_r(x)$.

Proof. Our proof diverges from Loeper's [26], but has the advantage of producing a radius $r_0 > 0$ independent of $t \in [0,1]$.

Let D and D^2 denote the gradient and the Hessian operators with respect to the fixed Euclidean coordinates. For any fixed $w \in T_x M$, Lemma 4.5 asserts

$$\frac{d^2}{dt^2}[D^2 f(t,x)](w,w) = \frac{1}{2} \sec_{(x,\bar x(t))}(w \oplus 0) \wedge (0 \oplus \dot{\bar x}(t)), \tag{B.2}$$

whence

$$\left| \frac{d^2}{dt^2}[D^2 f(t,x)](w,w) \right| \leq \frac{1}{2} n^2 |w|^2 |\dot{\bar x}(t)|^2 \sup_{i,\bar j,k,\bar l} |R_{i\bar j k \bar l}(x, \bar x(t))|.$$

from (2.2), and the Cauchy-Schwarz inequality $(\sum_{i=1}^n |w^i|)^2 \leq n|w|^2$.

On the other hand, setting $p^*(t) := Df(t,x) = -Dc(x, \bar x(t))$ yields the linear parameterization $p^*(t) = p^*(0) + tq^*$ as in (4.4). From this, it is easy to see that $w \oplus \dot{\bar x}(t)$ is an h-null vector at $(x, \bar x(t))$ if and only if $q^*_i w^i = 0$ where

$$q^*_i = \frac{dp^*_i}{dt}(t) = -c_{i\bar j}(x, \bar x(t)) \dot x^{\bar j}(t). \tag{B.3}$$

Note $q^* \neq 0$ because $\dot{\bar x}(0) \neq 0$. For any vector w in the nullspace of q^*, (B.1) and (B.2) imply

$$\frac{d^2}{dt^2}[D^2 f(t,x)](w,w) \geq C_0 |w|^2 |\dot{\bar x}(t)|^2.$$

To make these estimates independent of t, note (B.3) combines with nondegeneracy **(A2)** of the cost to imply that $\dot{\bar x}(t)$ and q^* have comparable Euclidean magnitudes

$$\frac{1}{\|D\bar D c\|_{C^0(N)}} |q^*| \leq |\dot{\bar x}(t)| \leq \|[D\bar D c]^{-1}\|_{C^0(N)} |q^*|. \tag{B.4}$$

It follows that

$$\begin{aligned} |\bar x(1) - \bar x(0)| &\leq \int_0^1 |\dot{\bar x}(\tau)| d\tau \\ &\leq \|[D\bar D c]^{-1}\|_{C^0(N)} |q^*|. \end{aligned} \tag{B.5}$$

Combining these estimates, we see

$$\text{(B.6)}\qquad \begin{cases} \left|\frac{d^2}{dt^2}[D^2 f(t,x)](w,w)\right| \leq C_1'|w|^2|q^*|^2 & \text{for general } w \\ \frac{d^2}{dt^2}[D^2 f(t,x)](w,w) \ \geq C_0'|w|^2|q^*|^2 & \text{for each } w \text{ with } q_i^* w^i = 0 \end{cases}$$

where

$$\text{(B.7)}\qquad C_0' = C_0/\|D\bar{D}c\|^2_{C^0(N)},$$

$$\text{(B.8)}\qquad C_1' = \frac{1}{2}n^2\|[D\bar{D}c]^{-1}\|^2_{C^0(N)} \sup_{i,\bar{j},k,\bar{l}} \|R_{i\bar{j}k\bar{l}}\|_{C^0(N)}.$$

In the following, we use (B.6) and a Taylor expansion argument. We work in a small ball $B_r(x) \subset M$ of x, whose radius r shall be determined in the course of the proof; note that the condition $B_r(x) \subset M$ is used only to have the Taylor expansion. Consider a thin cone

$$K_\theta := \{v \in T_x\Omega \mid \Big| -\cos\theta \leq \frac{q_i^* v^i}{|q^*||v|} \leq \cos\theta\}$$

around the nullspace of q^*. Here $\pi/4 < \theta < \pi/2$ will subsequently be chosen to be large, depending on C_0' and C_1'. We shall establish the theorem by taking $\theta \approx \pi/2$ in case 1 and $r > 0$ small in Case 2 below.

Case 1 (Second order): Let $v \in B_r(0) \cap K_\theta$. Then

$$\begin{aligned} &\max[f(0,x+v), f(1,x+v)] - f(t,x+v) \\ &\geq \ tf(1,x+v) + (1-t)f(0,x+v) - f(t,x+v) \\ &\geq [tD^2 f(1,x) + (1-t)D^2 f(0,x) - D^2 f(t,x)](v,v) - \|c\|_{C^3(N)}|v|^3 \end{aligned}$$

where the first order term $v(tp^*(0) + (1-t)p^*(1) - p^*(t))$ of the Taylor expansion has vanished due to the linearity of $p^*(t) = p^*(0) + tq^*$. To bound the second term using (B.6), rotate the Euclidean coordinates so that $v = (v_1, V)$, where $v_1 = q_i^* v^i/|q^*|$ and $(0, V)$ lies in the nullspace q^*. Note that for $v \in B_r(0) \cap K_\theta$ we have $|v_1| < |v|\cos\theta$ and hence $|v|^2 \sin^2\theta \leq |V|^2$. After the mean value theorem, (B.6) yields

$$\begin{aligned} &[tD^2 f(1,x) + (1-t)D^2 f(0,x) - D^2 f(t,x)](v,v) \\ &\geq \frac{1}{2}C_0' t(1-t)|q^*|^2|V|^2 - \frac{1}{2}C_1' t(1-t)|q^*|^2(2|v_1||V| + |v_1|^2) \\ &\geq \frac{1}{2}\Big[C_0'\sin^2\theta - 3C_1'\cos\theta\Big]t(1-t)|q^*|^2|v|^2 \\ &\geq \frac{1}{4}C_0' t(1-t)\left|\frac{\bar{x}(1) - \bar{x}(0)}{\|[D\bar{D}c]^{-1}\|_{C^0(N)}}\right|^2 |v|^2 \end{aligned}$$

noting (B.5), and provided we choose $\pi/4 < \theta < \pi/2$ sufficiently large that

$$\text{(B.9)}\qquad \sin^2\theta \geq \frac{1}{2} + \frac{3C_1'}{C_0'}\cos\theta.$$

This establishes the desired inequality for Case 1 for θ close to $\pi/2$.

Case 2 (First order): Let $v \in B_r(0) \setminus K_\theta$, and recall

$$\begin{aligned} p^*(1) - p^*(t) &= (1-t)q^* \\ p^*(0) - p^*(t) &= -tq^*. \end{aligned}$$

Now either $q_i^* v^i \le -|q^*||v|\cos\theta$ or $|q^*||v|\cos\theta \le q_i^* v^i$. If $q_i^* v^i \le 0$, Taylor expansion yields

$$\begin{aligned} &\max[f(0,x+v), f(1,x+v)] - f(t,x+v) \\ &\ge f(0,x+v) - f(t,x+v) \\ &= (p_i^*(0) - p_i^*(t))v^i + (D^2 f(0,x) - D^2 f(t,x))(v,v) + O(|v|^3) \\ &\ge t\cos\theta|v||q^*| - t\|D^2\bar{D}c\|_{C^0(N)} \max_{0<\tau<t}\{|\dot{\bar{x}}(\tau)|\}|v|^2 - \|c\|_{C^3(N)}|v|^3 \\ &\ge t|v||q^*|\Big(\cos\theta - \|c\|_{C^3(N)}\|[D\bar{D}c]^{-1}\|_{C^0(N)}|v|\Big) - \|c\|_{C^3(N)}|v|^3 \\ &\ge t|v||\bar{x}(1) - \bar{x}(0)|\Big(\|[D\bar{D}c]^{-1}\|^{-1}_{C^0(N)}\cos\theta - \|c\|_{C^3(N)}|v|\Big) - \|c\|_{C^3(N)}|v|^3 \end{aligned}$$

where (B.4)–(B.5) have been used. Similarly, if $q_i^* w^i \ge 0$, then

$$\begin{aligned} &\max[f(0,x+v), f(1,x+v)] - f(t,x+v) \\ &\ge (1-t)|v||\bar{x}(1) - \bar{x}(0)|\Big(\|[D\bar{D}c]^{-1}\|^{-1}_{C^0(N)}\cos\theta - \|c\|_{C^3(N)}|v|\Big) - \|c\|_{C^3(N)}|v|^3. \end{aligned}$$

If we assume

$$|v| < r \le r_0 := \frac{\|[D\bar{D}c]^{-1}\|^{-1}_{C^0(N)}\cos\theta}{\|c\|_{C^3(N)} + C_1 \operatorname{Diam}\bar{M}}, \tag{B.10}$$

then for $v \in B_r(0) \setminus K_\theta$ with either sign of $q_i^* v^i$ we find

$$\begin{aligned} &\max[f(0,x+v), f(1,x+v)] - f(t,x+v) \\ &\ge C_1 t(1-t)|v|^2|\bar{x}(1) - \bar{x}(0)|^2 - \|c\|_{C^3(N)}|v|^3. \end{aligned}$$

Combining Case 1 with Case 2 establishes the proposition; the constants have the desired dependence according to (B.7), (B.8), (B.9), (B.10), and (4.2). $\square$

Appendix C. On fitting sausages into balls

Loeper's Hölder continuity argument for optimal mappings $F : \Omega \longrightarrow \bar{\Omega}$ between domains $M = \Omega \subset \mathbf{R}^n$ and $\bar{M} = \bar{\Omega} \subset \mathbf{R}^n$ is based on a localized volume comparison. To set this up, for any optimal map which fails to be $C^{1/5}(\Omega; \operatorname{cl}\bar{\Omega})$, the next proposition identifies $\epsilon, \delta > 0$ and a null geodesic $t \in [0,1] \longrightarrow (x, \bar{x}(t)) \in \Omega \times \bar{\Omega}$, such that an ϵ-ball in coordinates around x and a δ-sausage $\mathcal{N}_\delta$ around the middle third of $\bar{x}(t)$ satisfy $F^{-1}(\mathcal{N}_\delta) \subset B_\epsilon(x)$. More precisely, we establish Proposition C.2 for the potential function u of the mapping $F = c\text{-Exp} \circ Du$.

We start by defining δ-neighbourhoods and relative δ-neighbourhoods. As usual, $[\bar{x}, \bar{y}]$ denotes the Euclidean convex hull of the two points $\bar{x}, \bar{y} \in \mathbf{R}^n$, and we write $Y \subset\subset \bar{\Omega}$ to indicate that the closure of Y is a compact subset of $\bar{\Omega}$.

Definition C.1. (Relative δ-neighbourhoods) *Given $Y \subset \bar\Omega \subset \mathbf{R}^n$, denote*

$$\mathcal{N}_\delta(Y) = \{\bar x \in \mathbf{R}^n \mid \text{there exists } \bar y \in Y \text{ such that } |\bar x - \bar y| < \delta\},$$

$$\mathcal{N}_\delta^{\bar\Omega}(Y) = \{x \in \bar\Omega \mid \text{there exists } \bar y \in Y \text{ such that } |\bar x - \bar y| < \delta \text{ and the line segment } [\bar x, \bar y] \subset \bar\Omega\}.$$

Note $\mathcal{N}_\delta^{\bar\Omega}(Y) = \bar\Omega \cap \mathcal{N}_\delta(Y)$ if $\bar\Omega \subset \mathbf{R}^n$ is convex. The point of this section is the following version of Loeper's Proposition 5.3 [26].

Proposition C.2. (Fitting sausages into balls) *Let $c \in C^4(\mathrm{cl}\, N)$ be strictly regular* **(A2)–(A3s)** *on the closure of the product $N = \Omega \times \bar\Omega$ of two bounded domains Ω and $\bar\Omega \subset \mathbf{R}^n$. Fix $u : \Omega \longrightarrow \mathbf{R}$ semiconvex with constant C, such that any mountain focused in $\mathrm{cl}\,\bar\Omega$ and supporting u to first-order is dominated by u throughout Ω. Choose a pair of points $(x_0, \bar x_0)$ and $(x_1, \bar x_1)$ from $\partial^c u \subset \mathrm{cl}(\Omega \times \bar\Omega)$ with $\bar x_0 \neq \bar x_1$ and $[x_0, x_1] \subset \Omega$. From the definition of $\partial^c u = \partial^c_{\mathrm{cl}\,\Omega} u$, both mountains*

$$f_i(\cdot) = -c(\cdot, \bar x_i) + c(x_i, \bar x_i) + u(x_i), \qquad i = 0, 1, \tag{C.1}$$

support u, hence there exists a point x where the valley of indifference $f_0(x) = f_1(x)$ intersects the line segment $[x_0, x_1] \subset \Omega$. Suppose a geodesic $t \in]0,1[\longrightarrow (x, \bar x(t)) \in N$ for the pseudo-metric (2.1) links $\bar x_0 = \lim_{t\to 0} \bar x(t)$ to $\bar x_1 = \lim_{t \to 1} \bar x(t)$. Then taking $C', C'', C''' > 0$ sufficiently small depending only on r_0 and C_1 from Proposition B.1, the semiconvexity constant C, $\|c\|_{C^3(N)}$, and $\mathrm{Diam}\,\bar\Omega$, implies: if $|x_1 - x_0| < (C'C'')^2 |\bar x_0 - \bar x_1|^5$ then for any $\epsilon > 0$ in the non-empty interval

$$\frac{1}{C'}\sqrt{\frac{|x_0 - x_1|}{|\bar x_0 - \bar x_1|}} \le \epsilon \le C''|\bar x_0 - \bar x_1|^2, \qquad \text{and} \tag{C.2}$$

$$\delta := C'''\,\epsilon\,|\bar x_0 - \bar x_1|^2, \tag{C.3}$$

such that $B_\epsilon(x) \subset\subset \Omega$ we have $[\mathrm{cl}\, B_\epsilon(x) \times \{\bar z\}] \cap \partial^c u$ non-empty for each $\bar z \in \mathcal{N}_\delta^{\bar\Omega}(\{\bar x(t) \mid \frac{1}{3} \le t \le \frac{2}{3}\})$.

Proof. We present a modified version of the original proof of Loeper [26].

In the course of the proof, $\delta > 0$ will be determined. Given $\bar z \in \mathcal{N}_\delta^{\bar\Omega}(\{\bar x(t), t \in [1/3, 2/3]\})$, define the mountain

$$g_{\bar z}(\cdot) := -c(\cdot, \bar z) + c(x, \bar z) + u(x)$$

focused at $\bar z$ which agrees with $u(\cdot)$ at x but need not be a supporting mountain. Our goal is to specify $\epsilon, \delta > 0$ with $B_\epsilon(x) \subset\subset \Omega$ such that

$$u(\cdot) \ge g_{\bar z}(\cdot) \quad \text{throughout } \partial B_\epsilon(x). \tag{C.4}$$

Once this is established, it implies the existence of z which realizes the global minimum of $u(\cdot) - g_{\bar z}(\cdot)$ on $\mathrm{cl}\, B_\epsilon(x)$. Adjusting the height of $f_{\bar z}(\cdot)$ by $u(z) - g_{\bar z}(z)$ yields a mountain which supports $u(\cdot)$ to first-order at z, hence globally by hypothesis. This implies $(z, \bar z) \in \partial^c u$ as desired.

By construction, $u \ge \max[f_0, f_1]$ on Ω. The choice of $\bar z$ implies $t \in [1/3, 2/3]$ exists with $|\bar z - \bar x(t)| \le \delta$ and

$$\text{the line segment } [\bar z, \bar x(t)] \subset \bar\Omega. \tag{C.5}$$

Assume $f_0(x) = 0 = f_1(x)$ without loss of generality, so

$$f_t(\cdot) := -c(\cdot, \bar{x}(t)) + c(x, \bar{x}(t)), \qquad t \in [0,1],$$

extends (C.1). Let $\epsilon < r_0$ be chosen to satisfy $B_\epsilon(x) \subset\subset \Omega$ later, where r_0 is from Proposition B.1. Then for each $y \in \partial B_\epsilon(x)$,

$$\begin{aligned} u(y) - g_{\bar{z}}(y) &\geq \max[f_0(y), f_1(y)] - g_{\bar{z}}(y) \\ &\geq f_t(y) - g_{\bar{z}}(y) + C_1 t(1-t)|\bar{x}_0 - \bar{x}_1|^2 |y-x|^2 - \|c(\cdot,\cdot)\|_{C^3(N)}|y-x|^3. \end{aligned} \tag{C.6}$$

To achieve (C.4) we want the right hand side nonnegative. First note

$$\begin{aligned} f_t(y) - g_{\bar{z}}(y) &= -c(y, \bar{x}(t)) + c(x, \bar{x}(t)) + c(y, \bar{z}) - c(x, \bar{z}) - u(x) \\ &\geq -\|c\|_{C^2(N)}|y-x||\bar{z} - \bar{x}(t)| - u(x) \end{aligned} \tag{C.7}$$

by second derivative estimates; this estimate is the only place where we use the condition (C.5) in the proof. Recall $x = (1-s)x_0 + sx_1$ for some $s \in [0,1]$. Semi-convexity implies

$$u(x) \leq (1-s)u(x_0) + su(x_1) + C|x_0 - x_1|^2/2$$

where

$$\begin{aligned} &(1-s)u(x_0) + su(x_1) \\ &= (1-s)f_0(x_0) + sf_1(x_1) \\ &\leq (1-s)Df_0(x)(x_0 - x) + sDf_1(x)(x_1 - x) + \frac{1}{2}\|c\|_{C^2(N)}(|x_0 - x|^2 + |x_1 - x|^2) \\ &\leq s(1-s)[Df_0(x) - Df_1(x)](x_0 - x_1) + \|c\|_{C^2(N)}|x_0 - x_1|^2/2 \\ &\leq s(1-s)\|D\bar{D}c\|_{C^0(N)}|\bar{x}_0 - \bar{x}_1||x_0 - x_1| + \|c\|_{C^2(N)}|x_1 - x_0|^2/2 \end{aligned}$$

Therefore,

$$u(x) \leq \|D\bar{D}c\|_{C^0(N)}|\bar{x}_0 - \bar{x}_1||x_0 - x_1| + (\|c\|_{C^2(N)} + C)|x_0 - x_1|^2/2.$$

From (C.6) and (C.7), the desired inequality (C.4) reduces to the following:

$$\begin{aligned} 0 \leq\ & C_1 t(1-t)|\bar{x}_0 - \bar{x}_1|^2\epsilon^2 - \|c\|_{C^2(N)}\epsilon\delta - \|c\|_{C^2(N)}|\bar{x}_0 - \bar{x}_1||x_0 - x_1| \\ & - (\|c\|_{C^2(N)} + C)|x_1 - x_0|^2/2 - \|c\|_{C^3(N)}\epsilon^3 \end{aligned}$$

which can be satisfied if the negative terms are small relative to the positive term $C_1 t(1-t)|\bar{x}_0 - \bar{x}_1|^2\epsilon^2 \gtrsim |\bar{x}_0 - \bar{x}_1|^2\epsilon^2$. This will be the case if

$$\begin{aligned} \delta &\lesssim |\bar{x}_0 - \bar{x}_1|^2\epsilon, \\ |x_0 - x_1| &\lesssim |\bar{x}_0 - \bar{x}_1|\epsilon^2, \\ \epsilon &\lesssim 1, \\ \epsilon &\lesssim |\bar{x}_0 - \bar{x}_1|^2 \end{aligned}$$

where all the relevant constants (for $\lesssim$) depend only on C, C_1, $\|c\|_{C^2(N)}$, and $\|c\|_{C^3(N)}$. Fixing suitable constants C', C'' and C''' implies the choices (C.2) and (C.3) suffice, provided C'' is taken small enough depending on r_0 and $\operatorname{Diam}\bar{\Omega}$ to have $\epsilon < r_0$ as required. For the interval (C.2) to be non-empty, we need

$|x_0 - x_1| < (C'C'')^2|\bar{x}_0 - \bar{x}_1|^5$, which is enough to ensure compatibility of the preceding four inequalities. □

Appendix D. Hölder continuity of optimal transport by volume comparison

Let us now show how the Hölder continuity of optimal mappings — Loeper's Theorem 3.4 and Proposition 5.5 [26] — follows from Theorem A.10, Proposition B.1, Proposition C.2, and Loeper's volume comparison argument, which we now recall. Note that Hölder continuity of the optimal transport map $F = c\text{-Exp} \circ Du$ implies Hölder differentiability of the corresponding potential $u \in \mathcal{S}^{-c}_{\mathrm{cl}\,\bar{M}}$.

Theorem D.1. (Hölder continuity [26]**)** *Assume $c \in C^4(\mathrm{cl}\,N)$ is bi-twisted and strictly regular* **(A1)–(A3s)** *on $\mathrm{cl}\,N$, where $N = \Omega \times \bar{\Omega} \subset \mathbf{R}^n \times \mathbf{R}^n$ is a bounded domain bi-convex with respect to the pseudo-metric* (2.1)*. Fix $m > 0$, and let ρ and $\bar{\rho}$ be probability measures on Ω and $\bar{\Omega}$ with Lebesgue densities $d\bar{\rho}/d\,\mathrm{vol} \geq m$ throughout $\bar{\Omega}$ and $d\rho/d\,\mathrm{vol} \in L^\infty(\Omega)$. Then there exists a map $F \in C^{1/\max\{5,4n-1\}}_{loc}(\Omega, \mathrm{cl}\,\bar{\Omega})$ between ρ and $\bar{\rho}$ which is optimal with respect to the transportation cost c.*

Proof. Note $c \in C^4(\mathrm{cl}\,N)$ since strict regularity extends to a neighbourhood of $\mathrm{cl}\,N$ by hypothesis. By Kantorovich duality, there exists $u \in \mathcal{S}^{-c}_{\mathrm{cl}\,\bar{\Omega}}$ such that the optimal γ in (1.1) vanishes outside $\partial^c u = \partial^c_{\mathrm{cl}\,\Omega} u$ [31] [40]. Moreover, both u and its transform $\bar{u} := u^{c^*}_{\mathrm{cl}\,\Omega}$ are globally Lipschitz and semiconvex from Lemma A.4 and our hypotheses on the cost. By Corollary A.5, $\mathrm{cl}\,\Omega \subset \mathrm{Dom}\,\partial^c u$. Since $u(x) + \bar{u}(\bar{x}) + c(x, \bar{x}) \geq 0$ with equality on $\partial^c u$, the bi-twist condition yields $\partial^c u \cap (\mathrm{Dom}\,Du \times \mathrm{Dom}\,\bar{D}\bar{u}) \subset \mathrm{Graph}(F) \cap \mathrm{Antigraph}(G)$ where $F := c\text{-Exp} \circ Du$ and $G := c^*\text{-Exp} \circ \bar{D}\bar{u}$ [18] [9] [27]. Here $\mathrm{Antigraph}(G) := \{(G(\bar{x}), \bar{x}) \mid \bar{x} \in \mathrm{Dom}\,\bar{D}\bar{u}\}$, and $\mathrm{Dom}\,Du \times \mathrm{Dom}\,\bar{D}\bar{u} \subset \Omega \times \bar{\Omega}$ is a set of full mass for $\rho \otimes \bar{\rho}$, a fortiori for γ. Moreover, $x = G(F(x))$ holds for ρ-a.e. $x \in \Omega$, and $\bar{x} = F(G(\bar{x}))$ for vol-a.e. $\bar{x} \in \bar{\Omega}$. We shall establish the theorem following [26].

Since N is horizontally-convex and **(A1′)** is satisfied, Theorem A.10 ensures the function u satisfies the hypotheses of Proposition C.2: namely that that any mountain supporting u to first-order supports u globally. Fix $\eta > 0$, and let $\Omega_\eta := \Omega \setminus \mathcal{N}_\eta(\partial\Omega)$. Then $x_0, x_1 \in \Omega_\eta$ with $|x_0 - x_1| < \eta/2$ implies $[x_0, x_1] \subset \Omega_{\eta/2}$, so any $\epsilon < \eta/2$ implies $\mathcal{N}_\epsilon([x_0, x_1]) \subset\subset \Omega$.

Pick distinct points $x_0, x_1 \in \Omega_\eta$ with $(x_i, \bar{x}_i) \in \partial^c u(x_i) \cap (\mathrm{Dom}\,Du \times \mathrm{Dom}\,\bar{D}\bar{u})$, $i = 0, 1$, such that

$$|x_0 - x_1| < \min\{(C'C'')^2|\bar{x}_0 - \bar{x}_1|^5, (\frac{\eta C'}{2})^2|\bar{x}_0 - \bar{x}_1|, \frac{\eta}{2}\}. \tag{D.1}$$

If there no such pair of points exists, then $F \in C^{1/5}(\Omega_\eta; \mathrm{cl}\,\bar{\Omega})$ with a Hölder constant depending on η, C', C'', and the Euclidean path diameter of Ω_η. Choosing x as in Proposition C.2, the vertical-convexity of N yields a geodesic $t \in [0, 1] \to (x, \bar{x}(t)) \in N$ connecting $(x, \bar{x}_0)$ to $(x, \bar{x}_1)$.

Unless $(x_0, \bar{x}_0)$ and $(x_1, \bar{x}_1)$ can be chosen to make $\epsilon = (|x_0 - x_1|/|\bar{x}_0 - \bar{x}_1|)^{1/2}/C'$ and hence $\delta = C'''\epsilon|\bar{x}_0 - \bar{x}_1|^2$ arbitrarily small in (C.2)–(C.3), the map F is Lipschitz

on Ω_η. Noting $0 < \epsilon < \eta/2$, Proposition C.2 yields

$$\mathcal{N}_\delta^{\bar\Omega}(\{\bar x(t), t \in [1/3, 2/3]\}) \subset \partial^c u(B_\epsilon(x)). \tag{D.2}$$

Since the cost c is bi-twisted, the map $G := c^*\text{-Exp} \circ D\bar u$ is well-defined on the set $\mathrm{Dom}\, D\bar u \subset \bar\Omega$ of differentiability for $\bar u := u^{c^*}_{\mathrm{cl}\,\Omega}$. This set has full Lebesgue measure (its complement has Hausdorff dimension $n-1$), and if $(x, \bar x) \in \partial^c u$ with $\bar x \in \mathrm{Dom}\, D\bar u$ then $x = G(\bar x)$. Now (D.2) asserts

$$\begin{aligned} B_\epsilon(x) &\supset G\Big(\mathcal{N}_\delta^{\bar\Omega}(\{\bar x(t), t \in [1/3, 2/3]\}) \cap \mathrm{Dom}\, D\bar u\Big) \cap \mathrm{Dom}\, Du \\ &= F^{-1}\Big(\mathcal{N}_\delta^{\bar\Omega}(\{\bar x(t), t \in [1/3, 2/3]\}) \cap \mathrm{Dom}\, D\bar u\Big) \cap \mathrm{Dom}\, Du. \end{aligned} \tag{D.3}$$

Since $F_\# \rho = \bar\rho$ has density $d\bar\rho / d\,\mathrm{vol} \geq m$, integrating ρ over these two sets yields

$$\|\rho\|_{L^\infty(\Omega)} \mathrm{vol}\, B_\epsilon(x) \geq m\, \mathrm{vol}(\mathcal{N}_\delta^{\bar\Omega}(\{\bar x(t), t \in [1/3, 2/3]\})). \tag{D.4}$$

For $c \in C^4(N)$ and twisted, c-Exp : $Dc(x, \bar\Omega) \longrightarrow \bar\Omega$ gives C^3 diffeomorphism. The pre-image of $\{\bar x(t)\}_{t \in [1/3, 2/3]}$ under this map is a segment in $Dc(x, \bar\Omega) \subset T^*\Omega$ according to Lemma 4.4. Noting that bi-convexity of N implies convexity of $Dc(x, \bar\Omega)$, it is not hard to see

$$\mathrm{vol}(\mathcal{N}_\delta^{\bar\Omega}(\{\bar x(t), t \in [1/3, 2/3]\})) \gtrsim \mathrm{vol}(\mathcal{N}_\delta(\{\bar x(t), t \in [1/3, 2/3]\})) \tag{D.5}$$

for $\delta > 0$ sufficiently small depending on c and N but not on $\bar x_0, \bar x_1$ or x. Recalling (C.3) then yields

$$\begin{aligned} \epsilon^n \|\rho\|_{L^\infty(\Omega)} &\gtrsim m\delta^{n-1} |\bar x_0 - \bar x_1| \\ &= m(C'''\epsilon)^{n-1} |\bar x_0 - \bar x_1|^{2n-1}. \end{aligned} \tag{D.6}$$

Thus whenever $\epsilon = (|x_0 - x_1| / |\bar x_0 - \bar x_1|)^{1/2} / C'$ and hence δ is sufficiently small we find

$$\|\rho\|^2_{L^\infty(\Omega)} |x_0 - x_1| \gtrsim (mC')^2 (C''')^{2n-2} |\bar x_0 - \bar x_1|^{4n-1}$$

so $F \in C^{1/(4n-1)} \cup C^{0,1}(\Omega_\eta, \mathrm{cl}\,\bar\Omega)$. Since $\eta > 0$ was arbitrary, the proof is complete. $\square$

Remark D.2 (Rougher measures). Note that no smoothness is required of the measures in the preceding theorem: $\bar\rho$ need not even be absolutely continuous with respect to Lebesgue and the support of ρ need not be connected. Replacing the hypothesis $d\rho / d\,\mathrm{vol} \in L^\infty(\Omega)$ by the existence of $p \in]n, +\infty]$ and $C(\rho) > 0$ such that

$$\rho(B_\epsilon(x)) \leq C(\rho) \epsilon^{n - n/p} \qquad \text{for all } \epsilon \geq 0, \quad x \in \Omega, \tag{D.7}$$

Loeper [26] was able to conclude $F \in C^{1/5}_{loc} \cup C^{(1-\frac{n}{p})/(4n-1-\frac{n}{p})}_{loc}(\Omega, \mathrm{cl}\,\bar\Omega)$ by assuming $n \geq 2$ and modifying the passage from (D.3) to (D.4) in the obvious way in the argument above. Replacing (D.7) by the even weaker condition that there exists $f : \mathbf{R}_+ \longrightarrow \mathbf{R}_+$ with $\lim_{\epsilon \to 0} f(\epsilon) = 0$ such that

$$\rho(B_\epsilon(x)) \leq f(\epsilon) \epsilon^{n-1} \qquad \text{for all } \epsilon \geq 0, \quad x \in \Omega, \tag{D.8}$$

Loeper was able to conclude continuity of $F : \Omega \longrightarrow \mathrm{cl}\,\bar{\Omega}$ similarly. In these cases ρ need not be absolutely continuous with respect to Lebesgue measure, though (D.8) precludes it from charging rectifiable sets of dimension less than or equal to $n-1$.

Remark D.3 (Hölder continuity of the boundary map). Strictly regular and bi-twisted on $\mathrm{cl}\,N$ means $N = \Omega \times \bar{\Omega}$ is compactly contained in a larger domain $N' \subset \mathbf{R}^n \times \mathbf{R}^n$ on which $c \in C^4(N')$ remains strictly regular and bi-twisted **(A0)**–**(A3s)**. If we allow $N' \not\subset \mathbf{R}^n \times \mathbf{R}^n$ to be a manifold which contains $\mathrm{cl}(N)$ compactly, we see the boundedness hypotheses on $M = \Omega$ and $\bar{M} = \bar{\Omega}$ become unnecessary in Propositions B.1–C.2 and Theorem D.1. If a bi-convex product domain $N' = \Omega' \times \bar{\Omega} \subset\subset \tilde{N}$ exists with $\Omega \subset\subset \Omega'$, then for $n \geq 2$ Theorem D.1 asserts $F \in C^{1/(4n-1)}_{loc}(\Omega', \mathrm{cl}\,\bar{\Omega})$, hence $F \in C^{1/(4n-1)}(\Omega, \mathrm{cl}\,\bar{\Omega})$.

Appendix E. Continuity of optimal transport on the sphere

This section presents a continuity result of optimal transport on the sphere as an *a priori* step toward Hölder continuity. It is known from the work of Cordero-Erausquin, McCann & Schmuckenschläger [13] that on a compact n-dimensional Riemannian manifold $(M = \bar{M}, g)$ with the Riemannian distance squared cost $c = d^2/2$ of Example 3.6, the optimal map stays away from cut locus for ρ-almost all points, when the measures ρ and $\bar{\rho}$ do not charge rectifiable sets of Hausdorff dimension $n-1$. On the round sphere, Delanoë & Loeper provided a quantitative estimate of the distance separating $F(x)$ from the cut locus of x in terms of $\|\log d\rho/d\,\mathrm{vol}\,\|_{L^\infty(\mathbf{S}^n)}$ and $\|\log d\bar{\rho}/d\,\mathrm{vol}\,\|_{L^\infty(\mathbf{S}^n)}$. In Corollary E.2 we use the preceding ideas to recover a form of this statement which, though less sharp than the result of [17], still suffices for our purpose: namely, to give a self-contained proof of Loeper's Hölder continuity result on the sphere, Theorem E.3. It is not known whether optimal maps with respect to strictly regular costs $c = d^2/2$ on other Riemannian manifolds stay uniformly away from the cut locus, however.

Theorem E.1. (Continuity of maps on the sphere) *Let $M = \partial \mathbf{B}_1(0) \subset \mathbf{R}^{n+1}$ be the Euclidean unit sphere, also denoted $\mathbf{S}^n$, equipped with either cost:*

(a) $c(x, \bar{x}) = \mathrm{dist}^2(x, \bar{x})/2$, *where* dist *denotes the great circle distance in* $\mathbf{S}^n$,
(b) $c(x, \bar{x}) = -\log|x - \bar{x}|$ *where* $|\cdot|$ *denotes the Euclidean distance of* $\mathbf{R}^{n+1}$.

Fix probability measures $\rho, \bar{\rho}$ on the round sphere $\mathbf{S}^n$, one satisfying (D.8) *and the other having a density $d\bar{\rho}/d\,\mathrm{vol} \geq m > 0$ with respect to Riemannian volume on $\mathbf{S}^n$. Then some continuous map F on $\mathbf{S}^n$ between ρ and $\bar{\rho}$ is optimal with respect to the cost function c and takes the form $F = c\text{-}Exp \circ Du$ with $u \in C^1(\mathbf{S}^n) \cap \mathcal{S}^{-c}_{\mathbf{S}^n}$.*

Proof. Let $N \subset \mathbf{S}^n \times \mathbf{S}^n$ denote the set of points where c is smooth, and $\Delta = \mathbf{S}^n \times \mathbf{S}^n \setminus N$ its complement, the singular set:

- for cost (a), $\Delta = \{(x, -x) \mid -x \text{ is the antipode of } x \in \mathbf{S}^n\}$
- for cost (b), $\Delta = \{(x, x) \mid x \in \mathbf{S}^n\}$.

Note the pseudo-Riemannian metric h of (2.1) is strictly regular on N, and N is bi-convex with respect to h. Moreover, this bi-convexity is 2-uniform (a fortiori strict)

in the sense that $Dc(x, \bar{N}(x)) \subset T_x^*\mathbf{S}^n$ is either (a) a disc of radius $\pi = \operatorname{Diam} \mathbf{S}^n$ or (b) the entire cotangent space.

The form $F = c\text{-Exp} \circ Du$ of the optimal map and Lipschitz potential $u \in \mathcal{S}^{-c}_{\mathbf{S}^n}$ can be found in (a) McCann [29], where u is semiconvex according to Cordero-Erausquin, McCann & Schmuckenschläger [13]. The same form can be deduced for cost (b) analogously, using the Lipschitz and semiconvexity from Proposition A.6. To derive a contradiction, suppose u is not differentiable at some point $x \in \mathbf{S}^n$. Since u is globally Lipschitz and semiconvex this means the compact convex set $\partial u(x) \subset T_x^*\mathbf{S}^n$ consists of more than one point. Choose two distinct co-vectors $p^* \neq q^*$ extremal in $\partial u(x)$. Extremality yields sequences $x_k \to x$ and $y_k \to x$ in $\operatorname{Dom} Du$ such that $Du(x_k) \to p^*$ and $Du(y_k) \to q^*$ by §25.6 [32]. Lemma A.9 asserts $(x_k, Du(x_k)) \in \operatorname{cl} \operatorname{Dom} c\text{-Exp}$ and provides a supporting mountain $-c(\cdot, \bar{x}_k) + c(x_k, \bar{x}_k) + u(x_k)$ for u at x_k which is focused at $\bar{x}_k = c\text{-Exp}'_{x_k} Du(x_k)$. The limit $k \to \infty$ yields a mountain supporting u at x and focused at $\bar{x} = c\text{-Exp}'_x p^*$; for the same reason, another mountain supporting u at x is focused at $\bar{y} = c\text{-Exp}'_x q^*$. Whether or not these mountains are distinct, a geodesic $t' \in]0,1[\longrightarrow (x, \bar{z}(t'))$ in $\{x\} \times \bar{N}(x)$ which extends from $(x, \bar{x})$ to $(x, \bar{y})$ is defined by $\bar{z}(t') := c\text{-Exp}_x[(1-t')p^* + t'q^*]$ according to Lemma 4.4. Applying Theorem 3.1 to (a) Example 3.7 and (b) Example 3.5 yields $\bar{z}(t') \in \partial^c u(x)$ for all $t' \in [0,1]$.

To exploit Proposition C.2, first localize the domains to subdomains of $\mathbf{R}^n$ in the following way. From the strict vertical convexity of $\mathbf{S}^n \times \mathbf{S}^n \setminus \Delta$, there is a subdomain $\bar{\Omega} \subset\subset \bar{N}(x)$ containing $\{\bar{z}(t')\}_{t' \in [1/4, 3/4]}$. Reparameterize this portion of the curve to get a geodesic $t \in [0,1] \longrightarrow (x, \bar{x}(t))$ with distinct endpoints $\bar{x}(0) = \bar{z}(1/4)$ and $\bar{x}(1) = \bar{z}(3/4)$. Choose a coordinate neighborhood $\Omega \subset \mathbf{S}^n$ of x which corresponds to a bounded domain $\Omega \subset\subset \mathbf{R}^n$; $\bar{\Omega}$ can be mapped to bounded domain in $\mathbf{R}^n$ by using its diffeomorphic preimage under $c\text{-Exp}_x$.

The rest of proof is an easy application of Proposition C.2 to this case $x_0 = x = x_1$ and $\bar{x}_0 = \bar{x}(0) \neq \bar{x}_1 = \bar{x}(1)$; we employ the notation of that proposition. Note that u satisfies the hypothesis of Proposition C.2 according to Theorem A.10.

We apply Proposition C.2 to $\operatorname{Graph} F \subset \partial^c u$. First notice from (C.2) that ϵ may be chosen arbitrarily in the range $]0, C''|\bar{x}_0 - \bar{x}_1|^2]$. To compare the volume of the ball $B_\epsilon(x)$ to the sausage $\mathcal{N}_\delta^{\bar{\Omega}}(\{\bar{x}(t), t \in [1/3, 2/3]\})$, we repeat steps (D.3)–(D.6) from the proof of Theorem D.1 to obtain

$$\begin{aligned} m(C'''\epsilon)^{n-1}|\bar{x}_0 - \bar{x}_1|^{2n-1} &= m\delta^{n-1}|\bar{x}_0 - \bar{x}_1| \\ &\lesssim \rho[B_\epsilon(x)] \\ &= o(\epsilon^{n-1}) \qquad \text{as } \epsilon \to 0, \end{aligned}$$

where the last equality is given by (D.8). The limit $\epsilon \to 0$ forces $\bar{x}_0 = \bar{x}_1$ — a contradiction! We therefore conclude u is differentiable, which implies $u \in C^1(\mathbf{S}^n)$ by the closedness of ∂u noted after (A.4). The continuity of $F = c\text{-Exp} \circ Du$ follows immediately. $\square$

Corollary E.2. (Uniformly Away from Singular set [17]**)** *Assume the hypotheses of Theorem E.1, namely* $(M = \mathbf{S}^n, \mathrm{dist}, \mathrm{vol})$ *is the round sphere and*

$F : \mathbf{S}^n \longrightarrow \mathbf{S}^n$ is an optimal map between a probability measure ρ satisfying (D.8) and one satisfying $d\bar\rho/d\,\mathrm{vol} \geq m > 0$ for the (a) Riemannian distance squared or (b) negative logarithm of the Euclidean cost. Let $N \subset \mathbf{S}^n \times \mathbf{S}^n$ be the set where c is smooth. There exists a constant $\delta = \delta(\rho, \bar\rho) > 0$ such that ρ-a.e. point $x \in \mathbf{S}^n$ satisfies

$$\mathrm{dist}(F(x), \mathbf{S}^n \setminus \bar N(x)) \geq \delta. \tag{E.1}$$

Proof. Since the map F is unique (up to ρ-negligible sets) [29], Theorem E.1 asserts it costs no generality to take $F = c\text{-Exp} \circ Du$ continuous on $\mathbf{S}^n$ and $u^{c^*c} = u \in C^1(\mathbf{S}^n)$. Recall for cost (a) $\bar N(x) = \{\bar x \in \mathbf{S}^n \mid \mathrm{dist}(x, \bar x) < \pi = \mathrm{Diam}_{\mathbf{S}^n} \mathbf{S}^n\}$, while for cost (b) $\bar N(x) = \{\bar x \in \mathbf{S}^n \mid \bar x \neq x\}$. In case (b) the desired estimate follows directly from the Lipschitz bound on Du provided by Proposition A.6, since $|Dc(x, \bar x)| \to \infty$ as $\bar x \to x$. If the corollary fails in case (a), there is a sequence of points $x_k \in \mathbf{S}^n$ with $\mathrm{dist}(F(x_k), \mathbf{S}^n \setminus \bar N(x_k)) \to 0$ as $k \to \infty$. Extracting a subsequential limit yields $x_k \to x \in \mathbf{S}^n$ with $F(x) \notin \bar N(x) = \mathrm{Ran}\, c\text{-Exp}_x$ antipodal to x. Due to the symmetry of the sphere, the convex set $\partial u(x)$ cannot then consist of a single point of non-zero magnitude; instead $\partial u(x) = \{p^* \in T^*_x\mathbf{S}^n \mid |p^*| \leq \pi\}$ contradicting the claim $u \in C^1(\mathbf{S}^n)$. □

Finally, we collect the ingredients above to conclude our direct proof of (a) Loeper's result [26], and (b) his analogous improvements of Caffarelli, Gutiérrez & Huang's result for the reflector antenna [8].

Theorem E.3. (Hölder continuity of optimal maps on the sphere [26], [8]**)** *As in Theorem E.1, assume ρ and $\bar\rho$ are probability measures on the round sphere $(M = \mathbf{S}^n, \mathrm{dist}, \mathrm{vol})$ satisfying (D.7) for some $p \in]n, \infty]$ and $d\bar\rho(x)/d\,\mathrm{vol} \geq m > 0$ for all $x \in \mathbf{S}^n$. For either cost (a) $c(x, \bar x) = \mathrm{dist}^2(x, \bar x)/2$ or (b) $c(x, \bar x) = -\log|x - \bar x|$ the optimal map F between ρ and $\bar\rho$ lies in $C^{1/5} \cup C^{(1-\frac{n}{p})/(4n-1-\frac{n}{p})}(\mathbf{S}^n, \mathbf{S}^n)$.*

Proof. We first set up a domain N to which we shall apply the proof of Theorem D.1. Let g_{ij} denote the standard metric tensor on the round sphere $\mathbf{S}^n$. For each $z \in \mathbf{S}^n$, let $-z$ denote its antipodal point. Observe that each g-geodesic ball $B_R(z)$ appears convex from $z \in \mathbf{S}^n$ for the Riemannian distance squared cost (a), and from $-z$ for the negative logarithmic cost (b); see Definition 2.5. In both cases, the convexity is 2-uniform in the sense that $Dc(z, B_R(\pm z))$ is a 2-uniform subset of $T_z\mathbf{S}^n$.

Let z be an arbitrary point in $\mathbf{S}^n$ and $\delta = \delta(\rho, \bar\rho)$ as in (E.1). Fix $\pi - \delta < R < \pi$. By compactness, we can choose $0 < r < \delta + R - \pi$ sufficiently small that the set $N = B_r(z) \times B_R(\pm z)$ remains vertically-convex, choosing the plus sign for cost (a) and the minus sign for cost (b). If we use g-geodesic polar coordinates we can regard the geodesic balls as Euclidean balls in the tangent spaces $T_z\mathbf{S}^n$ and $T_{\pm z}\mathbf{S}^n$.

We now apply the volume comparison argument of the proof of Theorem D.1 to the domain N. A few points should be clarified. Horizontal convexity of N was used in Theorem D.1 only to be able to apply Theorem A.10. But in the present context, we need not apply Theorem A.10 since we have its conclusion for u on the whole $\mathbf{S}^n$ directly by combining Theorem E.1 with Lemma A.9. Also F satisfies $\{(x, F(x)) \mid x \in B_r(z)\} \subset N$ by Corollary E.2. The remainder of the proof of Theorem D.1 goes

through in the present case, yielding $F \in C^{1/5}_{\rm loc} \cup C^{(1-\frac{n}{p})/(4n-1-\frac{n}{p})}_{\rm loc}(B_r(z), \mathbf{S}^n)$. This local estimate becomes global by compactness of $\mathbf{S}^n$, so the proof is complete. $\square$

References

[1] L.A. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lecture Notes in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[2] A. Besse. *Einstein Manifolds.* Springer-Verlag, Berlin, 1987.

[3] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

[4] L. Caffarelli. Allocation maps with general cost functions. In P. Marcellini et al, editor, *Partial Differential Equations and Applications*, number 177 in Lecture Notes in Pure and Appl. Math., pages 29–35. Dekker, New York, 1996.

[5] L.A. Caffarelli. The regularity of mappings with a convex potential. *J. Amer. Math. Soc.*, 5:99–104, 1992.

[6] L.A. Caffarelli. Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.*, 45:1141–1151, 1992.

[7] L.A. Caffarelli. Boundary regularity of maps with convex potentials — II. *Ann. of Math. (2)*, 144:453–496, 1996.

[8] L.A. Caffarelli, C. Gutiérrez, and Q. Huang. On the regularity of reflector antennas. *To appear in Ann. of Math. (2).*

[9] G. Carlier. Duality and existence for a class of mass transportation problems and economic applications. *Adv. Math. Econom.*, 5:1–21, 2003.

[10] P.-A. Chiappori, R.J. McCann, and L. Nesheim. Nonlinear hedonic pricing: finding equilibria through linear programming. *In preparation.*

[11] D. Cordero-Erausquin. Sur le transport de mesures périodiques. *C.R. Acad. Sci. Paris Sér. I Math.*, 329:199–202, 1999.

[12] D. Cordero-Erausquin. Non-smooth differential properties of optimal transport. In *Recent Advances in the Theory and Applications of Mass Transport*, Contemp. Math., pages 61–71. Amer. Math. Soc., Providence, 2004.

[13] D. Cordero-Erausquin, R.J. McCann and M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146:219–257, 2001.

[14] P. Delanoë. Classical solvability in dimension two of the second boundary-value problem associated with the Monge-Ampère operator. *Ann. Inst. H. Poincarè Anal. Non Linèaire*, 8:443–457, 1991.

[15] P. Delanoë. Gradient rearrangement for diffeomorphisms of a compact manifold. *Differential Geom. Appl.*, 20:145–165, 2004.

[16] P. Delanoë and Y. Ge. *In preparation.*

[17] P. Delanoë and G. Loeper. Gradient estimates for potentials of invertible gradient-mappings on the sphere. *Calc. Var. Partial Differential Equations*, 26:297–311, 2006.

[18] W. Gangbo. *Habilitation thesis.* Université de Metz, 1995.

[19] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.

[20] W. Gangbo and R.J. McCann. Shape recognition via Wasserstein distance. *Quart. Appl. Math.*, 58:705–737, 2000.

[21] T. Glimm and V. Oliker. Optical design of single reflector systems and the Monge-Kantorovich mass transfer problem. *J. Math. Sci.*, 117:4096–4108, 2003.

[22] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[23] Y.-H. Kim. Counterexamples to continuity of optimal transportation on positively curved riemannian manifolds. *Preprint.*

[24] Y.-H. Kim and R.J. McCann. *In preparation.*

[25] Y.-H. Kim and R.J. McCann. On the cost-subdifferentials of cost-convex functions. *Preprint at* arXiv 0706.1226v1.

[26] G. Loeper. On the regularity of maps solutions of optimal transportation problems. *Preprint at arXiv:math/0504138.*

[27] X.-N. Ma, N. Trudinger and X.-J. Wang. Regularity of potential functions of the optimal transportation problem. *Arch. Rational Mech. Anal.*, 177:151–183, 2005.

[28] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80:309–323, 1995.

[29] R.J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608, 2001.

[30] G. Monge. Mémoire sur la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704, 1781.

[31] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems.* Probab. Appl. Springer-Verlag, New York, 1998.

[32] R.T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1972.

[33] N.S. Trudinger. Recent developments in elliptic partial differential equations of Monge-Ampère type. In *International Congress of Mathematicians*, volume 3, pages 291–301. Eur. Math. Soc., Zürich, 2006.

[34] N.S. Trudinger and X.-J. Wang. On the second boundary value problem for Monge-Ampère type equations and optimal transportation. *Preprint at arXiv:math/0601086.*

[35] N.S. Trudinger and X.-J. Wang. On strict convexity and C^1-regularity of potential functions in optimal transportation. *Preprint at arXiv:math/0702807.*

[36] J. Urbas. On the second boundary value problem for equations of Monge-Ampère type. *J. Reine Angew. Math.*, 487:115–124, 1997.

[37] J.A. Viaclovsky. Conformal geometry and fully nonlinear equations. To appear in *World Scientific Memorial Volume for S.S. Chern.*

[38] J.A. Viaclovsky. Conformal geometry, contact geometry, and the calculus of variations. *Duke Math. J.*, 101:283–316, 2000.

[39] J.A. Viaclovsky. Conformally invariant Monge-Ampère equations: global solutions. *Trans. Amer. Math. Soc.*, 352:4371–4379, 2000.

[40] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, 2003.

[41] C. Villani. *Optimal Transport, Old and New.* St. Flour Lecture Notes. Springer, New York, to appear.

[42] X.-J. Wang. On the design of a reflector antenna. *Inverse Problems*, 12:351–375, 1996.

[43] X.-J. Wang. On the design of a reflector antenna II. *Calc. Var. Partial Differential Equations*, 20:329–341, 2004.

[44] M. Warren. Calibrations associated to Monge-Ampère equations. *Preprint at arXiv: math.AP/0702291v1.*

[45] J.G. Wolfson. Minimal Lagrangian diffeomorphisms and the Monge-Ampère equation. *J. Differential Geom.*, 46:335–373, 1997.

Department of Mathematics, University of Toronto, Toronto, Ontario Canada M5S 2E4

E-mail address: yhkim@math.toronto.edu, mccann@math.toronto.edu